(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]: **C12N**

(21) International Application Number: PCT/US00/17887

(22) International Filing Date: 28 June 2000 (28.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/349,015   7 July 1999 (07.07.1999)   US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US   09/349,015 (CIP)
Filed on   7 July 1999 (07.07.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **JONES, Karen, Ann** [GB/GB]; Cregneish, Elmdon, Saffron Walden, Essex CB11 4LT (GB). **VOLKMUTH, Wayne** [US/US]; 3620 El Encanto, Calabasas, CA 91302 (US). **WALKER,**

Michael, G. [CA/US]; 1050 Borregas Avenue, #80, Sunnyvale, CA 94089 (US).

(74) Agents: **MURRY, Lynn, E.** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

---

(54) Title: ATHEROSCLEROSIS-ASSOCIATED GENES

(57) **Abstract:** The invention provides novel atherosclerosis-associated polynucleotides and polypeptides encoded by those genes. The invention also provides expression vectors, host cells, and antibodies. The invention also provides methods for screening or purifying ligands and diagnosing or treating atherosclerosis.

ATHEROSCLEROSIS-ASSOCIATED GENES

TECHNICAL FIELD

The invention relates to 34 atherosclerosis-associated polynucleotides identified by their co-
5 expression with known atherosclerosis genes and their corresponding gene products. The invention
also relates to the use of these biomolecules in diagnosis, prognosis, prevention, treatment, and
evaluation of therapies for diseases associated with atherosclerosis.

BACKGROUND ART

Atherosclerosis is a disorder characterized by cellular changes in the arterial intima and the
10 formation of arterial plaques containing intra- and extracellular deposits of lipids. The resultant
thickening of artery walls and the narrowing of the arterial lumen is the underlying pathologic condition
in most cases of coronary artery disease, aortic aneurysm, peripheral vascular disease, and stroke. A
cascade of molecules is involved in the cellular morphogenesis, proliferation, and cellular migration
which results in an atherosclerotic lesion (Libby et al. (1997) Int J Cardiol 62:23-29).

15   A healthy artery consists of three layers. The vascular intima, lined by a monolayer of
endothelial cells in contact with the blood, contains smooth muscle cells in extracellular matrix. An
internal elastic lamina forms the border between the intima and the tunica media. The media contains
layers of smooth muscle cells surrounded by a collagen and elastin-rich extracellular matrix. An
external elastic lamina forms the border between the media and the adventitia. The adventitia contains
20 nerves and some mast cells and is the origin of the vasa vasorum which supplies blood to the outer
layers of the tunica media.

Initiation of an atherosclerotic lesion often occurs following vascular endothelial cell injury as a
result of hypertension, diabetes mellitus, hyperlipidemia, fluctuating shear stress, smoking, or
transplant rejection. The injury results in the local release of nitric oxide and superoxide anions which
25 react to form cytodestructive peroxynitrite radicals, causing injury to the endothelium and myocytes of
the intima. This cellular injury leads to the expression of a variety of molecules that produce local and
systemic effects. The initial cellular response to injury includes the release of mediators of
inflammation such as cytokines, complement components, prostaglandins, and downstream
transcription factors. These molecules promote monocyte infiltration of the vascular intima and lead to
30 the upregulation of adhesion molecules which encourage attachment of the monocytes to the damaged
endothelial cells. Additionally, components of the extracellular matrix including collagens,
fibrinogens, and matrix Gla protein are induced and provide sites for monocyte attachment. Annexins,
plasminogen activator inhibitor 1, and nitric oxide synthases are triggered to counteract these effects.

Monocytes that infiltrate the lesion accumulate modified low density lipoprotein lipid through
35 scavenger receptors such as CD36 and macrophage scavenger receptor type I. The abundance of
modified lipids is a factor in atherogenesis and is influenced by modifying enzymes such as lipoprotein

lipase, carboxyl ester lipase, serum amyloid P component, LDL-receptor related protein, microsomal triglyceride transfer protein, and serum esterases such as paraoxonase. Lipid metabolism is governed by cholesterol biosynthesis enzymes such as 3-hydroxy-3-methylglutaryl coenzyme A synthase, and products of the apolipoprotein genes. Modified lipid stabilization and accumulation is aided by

5   perilipin and alpha-2-macroglobulin.

As monocytes accumulate in the lesion, they can rupture and release free cholesterol, cytokines, and procoagulants into the surrounding environment. This process leads to the development of a plaque which consists of a mass of lipid-engorged monocytes and a lipid-rich necrotic core covered by a fibrous cap. The gradual progression of plaque growth is punctuated by thrombus formation which

10 leads to clinical symptoms such as unstable angina, myocardial infarction, or stroke. Thrombus formation is initiated by episodic plaque rupture which exposes flowing blood to tissue factors, which induce coagulation, and collagen, which activates platelets. After initiation of the atherosclerotic lesion, enzymes that degrade extracellular matrix components such as matrix metalloproteinases and cathepsin K are up-regulated, and their inhibitors are down-regulated. This results in destabilization of

15 the atherosclerotic lesion and subsequent complications including myocardial infarction, angina, and stroke. Further arterial occlusion and infiltration increase with the expression of coagulation factors and down-regulation of their inhibitors, antithrombin III, and lipoprotein-associated coagulation inhibitor.

Smooth muscle cells build up in the arterial media and constitute one of the principal cell types

20 in atherosclerotic and restenotic lesions. They show a high degree of plasticity and are able to shift between a differentiated, contractile phenotype and a less differentiated, synthetic phenotype. This modulation occurs as a response to factors secreted from cells at the site of vascular injury and results in structural reorganization with a loss of myofilaments and the formation of an extensive endoplasmic reticulum and a large Golgi complex. Genes encoding secreted protein, acidic and rich in cysteine

25 (SPARC) and endothelin-1 contribute to these changes. At the same time, the expression of cytoskeletal proteins such as calponin, myosin, desmin, and other gene products in the cells is altered. As a result, the smooth muscle cells lose their contractility and become able to migrate from the media to the intima, to proliferate, and to secrete extracellular matrix components which contribute to arterial intimal thickening.

30   The initiation and progression of atherosclerotic lesion development requires the interplay of various molecular pathways Many genes that participate in these processes are known, and some of them have been shown to have a direct role in atherosclerosis pathogenesis by animal model experiments, in vitro assays, and epidemiological studies (Krettek et al. (1997) Arterioscler Thromb Vasc Biol 17:2897-2903; Fisher et al. (1997) Atherosclerosis 135:145-159; Shih et al. (1998)

35 Circulation 95:2684-2693; and Bocan et al. (1998) Atherosclerosis 139:21-30).

The present invention satisfies a need in the art by providing new compositions that are useful for diagnosis, prognosis, treatment, prevention, and evaluation of therapies for diseases associated with atherosclerosis. We have implemented a method for analyzing gene expression patterns and have identified 34 atherosclerosis-associated polynucleotides through their co-expression with 66 known

5 atherosclerosis-associated genes.

SUMMARY OF THE INVENTION

The invention provides for a substantially purified polynucleotide comprising a gene that is coexpressed with one or more known atherosclerosis-associated genes in a biological sample. Known atherosclerosis-associated genes include and encode human 22kDa smooth muscle protein, calponin,

10 desmin, smooth muscle myosin heavy chain, alpha tropomyosin, human tissue inhibitor of metalloproteinase 3, human tissue inhibitor of metalloproteinase-2, human tissue inhibitor of metalloproteinase-4, pro alpha 1(I) collagen, collagen alpha-2 type I, collagen alpha-6 type I, procollagen alpha 2(V), collagen VI alpha-2, type VI collagen alpha3, pro-alpha-1 type 3 collagen, pro-alpha-1 (V) collagen, collagenase type IV/ matrix metalloproteinase 9/gelatinase B, matrix Gla protein,

15 cathepsin K, fibrinogen beta chain gene, fibrinogen gamma chain gene, pre-pro-von Willebrand factor, coagulation factor II/ prothrombin, coagulation factor XII, coagulation factor VII, platelet endothelial cell adhesion molecule, lipoprotein-associated coagulation inhibitor, antithrombin III variant, plasminogen activator inhibitor-1, lipoprotein lipase, alpha-2-macroglobulin, apolipoprotein AI, apolipoprotein AII, apolipoprotein B-100, lipoprotein apoCII, pre-apolipoprotein CIII, apolipoprotein

20 apo C-IV, macrophage scavenger receptor type I, human antigen CD36 gene, serum amyloid P component, carboxyl ester lipase gene, paraoxonase 1, paraoxonase 2, paraoxonase 3, LDL-receptor related protein, hepatic triglyceride lipase, 3-hydroxy-3-methylglutaryl coenzyme A synthase, very low density lipoprotein receptor, microsomal triglyceride transfer protein, perilipin, endothelin-1, endothelin receptor A, interleukin 6, interleukin 1, complement protein C8 alpha, complement component C9,

25 prostaglandin D2 synthase, annexin II/lipocortinII, annexin I/lipocortin, prostaglandin-endoperoxide synthase 2, insulin-like growth factor binding protein-1, secreted protein, acidic and rich in cysteine, human NF-kappa-B transcription factor, angiotensinogen, nitric oxide synthase 3, and nitric oxide synthase 2.

The invention also provides a substantially purified polynucleotide comprising a gene that is

30 coexpressed with one or more known atherosclerosis-associated genes in a plurality of samples. In one aspect, the polynucleotide comprises a polynucleotide sequence selected from a polynucleotide encoding a peptide selected from SEQ ID NOs:1-34; a polynucleotide sequence complementary to the polynucleotide sequence of SEQ ID NOs:1-34; and a probe comprising at least 18 sequential nucleotides of the polynucleotide sequence of SEQ ID NOs:1-34 or their complements. The invention

35 further provides a pharmaceutical composition comprising a polynucleotide and a pharmaceutical;

carrier.

The invention additionally provides methods for using a polynucleotide. One method uses the
polynucleotide to screen a library of molecules or compounds to identify at least one ligand which
specifically binds the polynucleotide and comprises combining the polynucleotide with a library of

5   molecules or compounds under conditions to allow specific binding and detecting specific binding,
thereby identifying a ligand which specifically binds the polynucleotide. In this first method, the
library is selected from DNA molecules, RNA molecules, PNAs, mimetics, and proteins; and the ligand
identified using the method may be used to modulate the activity of the polynucleotide. A second
method uses the polynucleotide to purify a ligand which specifically binds the polynucleotide and

10   comprises combining the polynucleotide with a sample under conditions to allow specific binding,
detecting specific binding between the polynucleotide and a ligand, recovering the bound
polynucleotide, and separating the polynucleotide from the ligand, thereby obtaining purified ligand. A
third method uses the polynucleotide to diagnose a disease or condition associated with the altered
expression of a gene that is coexpressed with one or more known atherosclerosis-associated genes in a

15   plurality of biological samples and comprises hybridizing a polynucleotide to a sample under
conditions to form one or more hybridization complexes, detecting the hybridization complexes, and
comparing the levels of the hybridization complexes with the level of hybridization complexes in a
non-diseased sample, wherein the altered level of hybridization complexes compared with the level of
hybridization complexes of a non-diseased sample indicates the presence of the disease or condition.

20   A fourth method uses the polynucleotide to produce a polypeptide and comprises culturing a host cell
containing an expression vector containing the polynucleotide under conditions for expression of the
polypeptide and recovering the polypeptide from cell culture.

The invention provides a substantially purified polypeptide comprising the product of a gene
that is coexpressed with one or more known atherosclerosis-associated genes in a plurality of samples.

25   The invention also provides a polypeptide comprising a polypeptide sequence selected from the
polypeptides encoded by SEQ ID NOs:1-34 and an oligopeptide sequence comprising at least 6
sequential amino acids of the polypeptide sequence of encoded by SEQ ID NOs:1-34. The further
provides a polypeptide comprising the amino acid sequence of SEQ ID NO:35. The invention still
further provides a pharmaceutical composition comprising a polypeptide and a pharmaceutical carrier.

30   The invention additionally provides methods for using a polypeptide. One method uses the
polypeptide to screen a library of molecules or compounds to identify at least one ligand which
specifically binds the polypeptide and comprises combining the polypeptide with the library of
molecules or compounds under conditions to allow specific binding and detecting specific binding
between the polypeptide and ligand, thereby identifying a ligand which specifically binds the

35   polypeptide. In this method, the library is selected from DNA molecules, RNA molecules, PNAs,

mimetics, polypeptides, agonists, antagonists, and antibodies; and the ligand identified using the method is used to modulate the activity of the polypeptide. A second method uses the polypeptide to purify a ligand from a sample and comprises combining the polypeptide with a sample under conditions to allow specific binding, detecting specific binding between the polypeptide and a ligand, recovering

5 the bound polypeptide, and separating the polypeptide from the ligand, thereby obtaining purified ligand. A third method uses the polypeptide to treat or to prevent a disease associated with the altered expression of a gene that is coexpressed with one or more known atherosclerosis-associated genes in a subject in need and comprises administering to the subject in need the pharmaceutical composition containing the polypeptide in an amount effective for treating the disease.

10   The invention provides an antibody or Fab comprising an antigen binding site, wherein the antigen binding site specifically binds to the polypeptide. The invention also provides a method for treating a disease associated with the altered expression of a gene that is coexpressed with one or more known atherosclerosis-associated genes in a subject in need, the method comprising the step of administering to the subject in need the antibody or the Fab in an amount effective for treating the

15 disease. The invention further provides an immunoconjugate comprising the antigen binding site of the antibody or Fab joined to a therapeutic agent. The invention additionally provides a method for treating a disease associated with the altered expression of a gene that is coexpressed with one or more known atherosclerosis-associated genes in a subject in need, the method comprising the step of administering to the subject in need the immunoconjugate in an amount effective for treating the disease.

20   **BRIEF DESCRIPTION OF THE SEQUENCE LISTING**

          The Sequence Listing provides exemplary atherosclerosis-associated gene sequences including polynucleotide sequences SEQ ID NOs:1-34 and the polypeptide sequence, SEQ ID NO:35. Each sequence is identified by a sequence identification number (SEQ ID NO).

                           DESCRIPTION OF THE INVENTION

25   It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include the plural reference unless the context clearly dictates otherwise. For example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

30 **Definitions**

          "Atherosclerosis-associated gene" refers to a gene or polynucleotide that exhibits a statistically significant coexpression pattern with known atherosclerosis-associated genes which are useful in the diagnosis, treatment, prognosis, or prevention of atherosclerosis.

          "Known atherosclerosis-associated gene" refers to a sequence which has been previously

35 identified as useful in the diagnosis, treatment, prognosis, or prevention of atherosclerosis and includes

polynucleotides encoding human 22kDa smooth muscle protein, calponin, desmin, smooth muscle myosin heavy chain, alpha tropomyosin, human tissue inhibitor of metalloproteinase 3, human tissue inhibitor of metalloproteinase-2, human tissue inhibitor of metalloproteinase-4, pro alpha 1(I) collagen, collagen alpha-2 type I, collagen alpha-6 type I, procollagen alpha 2(V), collagen VI alpha-2, type VI

5   collagen alpha3, pro-alpha-1 type 3 collagen, pro-alpha-1 (V) collagen, collagenase type IV/ matrix metalloproteinase 9/gelatinase B, matrix Gla protein, cathepsin K, fibrinogen beta chain gene, fibrinogen gamma chain gene, pre-pro-von Willebrand factor, coagulation factor II/prothrombin, coagulation factor XII, coagulation factor VII, platelet endothelial cell adhesion molecule, lipoprotein-associated coagulation inhibitor, antithrombin III variant, plasminogen activator inhibitor-1, lipoprotein

10   lipase, alpha-2-macroglobulin, apolipoprotein AI, apolipoprotein AII, apolipoprotein B-100, lipoprotein apoCII, pre-apolipoprotein CIII, apolipoprotein apo C-IV, macrophage scavenger receptor type I, human antigen CD36 gene, serum amyloid P component, carboxyl ester lipase gene, paraoxonase 1, paraoxonase 2, paraoxonase 3, LDL-receptor related protein, hepatic triglyceride lipase, 3-hydroxy-3-methylglutaryl coenzyme A synthase, very low density lipoprotein receptor, microsomal triglyceride

15   transfer protein, perilipin, endothelin-1, endothelin receptor A, interleukin 6, interleukin 1, complement protein C8 alpha, complement component C9, prostaglandin D2 synthase, annexin II/lipocortinII, annexin I/lipocortin, prostaglandin-endoperoxide synthase 2, insulin-like growth factor binding protein-1, secreted protein, acidic and rich in cysteine, human NF-kappa-B transcription factor, angiotensinogen, nitric oxide synthase 3, and nitric oxide synthase 2. Typically, this means that the

20   known gene is expressed at higher levels (i.e., has more abundant transcripts) in atherosclerotic lesions than in normal or non-diseased arterial intima or any other tissue.

"Ligand" refers to any molecule, agent, or compound which will bind specifically to a complementary site on a polynucleotide or polypeptide. Such ligands stabilize or modulate the activity of polynucleotides or polypeptides of the invention. For example, ligands are libraries of inorganic and

25   organic molecules or compounds such as nucleic acids, proteins, peptides, carbohydrates, fats, and lipids.

"NSEQ" refers generally to a polynucleotide sequence of the present invention, including SEQ ID NO:1-34. "PSEQ" refers generally to a polypeptide sequence of the present invention, including SEQ ID NO:35.

30   A "fragment" refers to a nucleic acid sequence that is preferably at least 20 nucleotides in length, more preferably 40 nucleotides, and most preferably 60 nucleotides in length, and encompasses, for example, fragments consisting of 1-50, 51-400, 401-4000, 4001-12,000 nucleotides, and the like, of SEQ ID NO:1-34.

"Gene" refers to the partial or complete coding sequence of a gene including 5' or 3'

35   untranslated regions. The gene may be in a sense or antisense (complementary) orientation.

"Polynucleotide" refers to a nucleic acid, nucleic acid sequence, oligonucleotide, nucleotide, or any fragment thereof. It may be DNA or RNA of genomic or synthetic origin, double-stranded or single-stranded, and combined with carbohydrate, lipids, protein or other materials to perform a particular activity or form a useful composition. "Oligonucleotide" is substantially equivalent to the

5   terms amplimer, primer, oligomer, element, and probe.

"Polypeptide" refers to an amino acid, amino acid sequence, oligopeptide, peptide, or protein or portions thereof whether naturally occurring or synthetic.

A "portion" refers to peptide sequence which is preferably at least 5 to about 15 amino acids in length, most preferably at least 10 amino acids long, and which retains some biological or

10   immunological activity of, for example, a portion of SEQ ID NO:35.

"Sample" is used in its broadest sense. A sample containing nucleic acids may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate; a cell; a tissue; a tissue print; and the like.

"Substantially purified" refers to a nucleic acid or an amino acid sequence that is removed from

15   its natural environment and that is isolated or separated, and is at least about 60% free, preferably about 75% free, and most preferably about 90% free, from other components with which it is naturally present.

"Substrate" refers to any rigid or semi-rigid support to which polynucleotides or polypeptides are bound and includes membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic

20   beads, gels, capillaries or other tubing, plates, polymers, and microparticles with a variety of surface forms including wells, trenches, pins, channels and pores.

A " variant" refers to a polynucleotide or polypeptide whose sequence diverges from SEQ ID NO:1-35. Polynucleotide sequence divergence may result from mutational changes such as deletions, additions, and substitutions of one or more nucleotides; it may also be introduced to accommodate

25   differences in codon usage. Each of these types of changes may occur alone, or in combination, one or more times in a given sequence.

THE INVENTION

The present invention encompasses a method for identifying biomolecules that are associated with a specific disease, regulatory pathway, subcellular compartment, cell type, tissue type, or species.

30   In particular, the method identifies polynucleotides useful in diagnosis, prognosis, treatment, prevention, and evaluation of therapies for diseases associated with atherosclerosis including, but not limited to, stroke, myocardial infarction, hypertension, transient cerebral ischemia, mesenteric ischemia, coronary artery disease, angina pectoris, peripheral vascular disease, intermittent claudication, renal artery stenosis, and hypertension.

35   The method entails first identifying polynucleotides that are expressed in a plurality of cDNA

7

libraries. The identified polynucleotides include genes of known or unknown function which are expressed in a specific disease process, subcellular compartment, cell type, tissue type, or species. The expression patterns of the genes with known function are compared with those of genes with unknown function to determine whether a specified coexpression probability threshold is met. Through this comparison, a subset of the polynucleotides having a high coexpression probability with the known genes can be identified. The high coexpression probability correlates with a particular coexpression probability threshold which is preferably less than 0.001 and more preferably less than 0.00001.

The polynucleotides originate from cDNA libraries derived from a variety of sources including, but not limited to, eukaryotes such as human, mouse, rat, dog, monkey, plant, and yeast; prokaryotes such as bacteria; and viruses. These polynucleotides can also be selected from a variety of sequence types including, but not limited to, expressed sequence tags (ESTs), assembled polynucleotide sequences, full length gene coding regions, promoters, introns, enhancers, 5' untranslated regions, and 3' untranslated regions. To have statistically significant analytical results, the polynucleotides need to be expressed in at least three cDNA libraries.

The cDNA libraries used in the coexpression analysis of the present invention can be obtained from adrenal gland, biliary tract, bladder, blood cells, blood vessels, bone marrow, brain, bronchus, cartilage, chromaffin system, colon, connective tissue, cultured cells, embryonic stem cells, endocrine glands, epithelium, esophagus, fetus, ganglia, heart, hypothalamus, immune system, intestine, islets of Langerhans, kidney, larynx, liver, lung, lymph, muscles, neurons, ovary, pancreas, penis, peripheral nervous system, phagocytes, pituitary, placenta, pleurus, prostate, salivary glands, seminal vesicles, skeleton, spleen, stomach, testis, thymus, tongue, ureter, uterus, and the like. The number of cDNA libraries selected can range from as few as 3 to greater than 10,000. Preferably, the number of the cDNA libraries is greater than 500.

In a preferred embodiment, genes are assembled from related sequences, such as assembled sequence fragments derived from a single transcript. Assembly of the sequences can be performed using sequences of various types including, but not limited to, ESTs, extensions, or shotgun sequences. In a most preferred embodiment, the polynucleotide sequences are derived from human sequences that have been assembled using the algorithm disclosed in "Database and System for Storing, Comparing and Displaying Related Biomolecular Sequence Information", Lincoln et al. Serial No:60/079,469, filed March 26, 1998, incorporated herein by reference.

Experimentally, differential expression of the polynucleotides can be evaluated by methods including, but not limited to, differential display by spatial immobilization or by gel electrophoresis, genome mismatch scanning, representational difference analysis, and transcript imaging. Additionally, differential expression can be assessed by microarray technology. These methods may be used alone or in combination.

Known atherosclerosis-associated genes are selected based on the use of these genes as diagnostic or prognostic markers or as therapeutic targets.

The procedure for identifying novel genes that exhibit a statistically significant coexpression pattern with known atherosclerosis-associated genes is as follows. First, the presence or absence of a

5   gene in a cDNA library is defined: a gene is present in a cDNA library when at least one cDNA fragment corresponding to that gene is detected in a cDNA sample taken from the library, and a gene is absent from a library when no corresponding cDNA fragment is detected in the sample.

Second, the significance of gene coexpression is evaluated using a probability method to measure a due-to-chance probability of the coexpression. The probability method can be the Fisher

10 exact test, the chi-squared test, or the kappa test. These tests and examples of their applications are well known in the art and can be found in standard statistics texts (Agresti (1990) Categorical Data Analysis, John Wiley & Sons, New York NY; Rice (1988) Mathematical Statistics and Data Analysis, Duxbury Press, Pacific Grove CA). A Bonferroni correction (Rice, supra, p. 384) can also be applied in combination with one of the probability methods for correcting statistical results of one gene versus

15 multiple other genes. In a preferred embodiment, the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set preferably to less than 0.001, more preferably to less than 0.00001.   To determine whether two genes, A and B, have similar coexpression patterns, occurrence data vectors can be generated as illustrated in Table 1. The presence of a gene occurring at least once in a library is indicated by a one, and its absence from the library, by a

20 zero.

Table 1. Occurrence data for genes A and B

	Library 1	Library 2	Library 3	...	Library N
gene A	1	1	0	...	0
gene B	1	0	1	...	0

25

For a given pair of genes, the occurrence data in Table 1 can be summarized in a 2 x 2 contingency table.

Table 2. Contingency table for co-occurrences of genes A and B

	Gene A present	Gene A absent	Total
Gene B present	8	2	10
Gene B absent	2	18	20
Total	10	20	30

Table 2 presents co-occurrence data for gene A and gene B in a total of 30 libraries. Both gene

35 A and gene B occur 10 times in the libraries. Table 2 summarizes and presents: 1) the number of times

gene A and B are both present in a library; 2) the number of times gene A and B are both absent in a library; 3) the number of times gene A is present, and gene B is absent; and 4) the number of times gene B is present, and gene A is absent. The upper left entry is the number of times the two genes co-occur in a library, and the middle right entry is the number of times neither gene occurs in a library.

5   The off diagonal entries are the number of times one gene occurs, and the other does not. Both A and B are present eight times and absent 18 times. Gene A is present, and gene B is absent, two times; and gene B is present, and gene A is absent, two times. The probability ("p-value") that the above association occurs due to chance as calculated using a Fisher exact test is 0.0003. Associations are generally considered significant if a p-value is less than 0.01 (Agresti, supra; Rice, supra).

10   This method of estimating the probability for coexpression of two genes makes several assumptions. The method assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent, because more than one library may be obtained from a single subject or tissue. Nor are they entirely identically sampled, because different numbers of cDNAs may be sequenced from each library. The number of

15   cDNAs sequenced typically ranges from 5,000 to 10,000 cDNAs per library. In addition, because a Fisher exact coexpression probability is calculated for each gene versus 45,233 other assembled genes, a Bonferroni correction for multiple statistical tests is used.

The present invention identifies 34 novel atherosclerosis-associated polynucleotides that exhibit strong association with genes known to be specific to atherosclerosis. The results presented in Table 4

20   show that the expression of the 34 novel atherosclerosis-associated polynucleotides has direct or indirect association with the expression of known atherosclerosis-associated genes. Therefore, the novel atherosclerosis-associated polynucleotides can potentially be used in diagnosis, treatment, prognosis, or prevention of diseases associated with atherosclerosis or in the evaluation of therapies for atherosclerosis. Further, the gene products of the 34 novel atherosclerosis-associated polynucleotides

25   are either potential therapeutics or targets of therapeutics against atherosclerosis.

Therefore, in one embodiment, the present invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:1-34. These 34 polynucleotides are shown by the method of the present invention to have strong coexpression association with known atherosclerosis-associated genes and with each other. The invention also encompasses a variant of the polynucleotide sequence,

30   its complement, or 18 consecutive nucleotides of a sequence provided in the above described sequences. Variant polynucleotide sequences typically have at least about 75%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to NSEQ.

NSEQ or the encoded PSEQ may be used to search against the GenBank primate (pri), rodent (rod), mammalian (mam), vertebrate (vrtp), and eukaryote (eukp) databases, SwissProt, BLOCKS

35   (Bairoch et al. (1997) Nucleic Acids Res 25:217-221), PFAM, and other databases that contain

previously identified and annotated motifs, sequences, and gene functions. Methods that search for

primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) Prot Eng 5:35-51)

as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul (1993) J Mol Evol

36:290-300; Altschul et al. (1990) J Mol Biol 215:403-410), BLOCKS (Henikoff and Henikoff (1991)

5   Nucleic Acids Res 19:6565-6572), Hidden Markov Models (HMM; Eddy (1996) Cur Opin Str Biol

6:361-365; Sonnhammer et al. (1997) Proteins 28:405-420), and the like, can be used to manipulate and

analyze nucleotide and amino acid sequences. These databases, algorithms and other methods are well

known in the art and are described in Ausubel et al. (1997; Short Protocols in Molecular Biology, John

Wiley & Sons, New York NY, unit 7.7) and in Meyers (1995; Molecular Biology and Biotechnology,

10   Wiley VCH, New York NY, p 856-853).

        Also encompassed by the invention are polynucleotide sequences that are capable of

hybridizing to SEQ ID NO:1-34, and fragments thereof under stringent conditions. Stringent

conditions can be defined by salt concentration, temperature, and other chemicals and conditions well

known in the art. Conditions can be selected, for example, by varying the concentrations of salt in the

15   prehybridization, hybridization, and wash solutions or by varying the hybridization and wash

temperatures. With some substrates, the temperature can be decreased by adding formamide to the

prehybridization and hybridization solutions.

        Hybridization can be performed at low stringency, with buffers such as 5xSSC with 1% sodium

dodecyl sulfate (SDS) at 60°C, which permits complex formation between two nucleic acid sequences

20   that contain some mismatches. Subsequent washes are performed at higher stringency with buffers

such as 0.2xSSC with 0.1% SDS at either 45°C (medium stringency) or 68°C (high stringency), to

maintain hybridization of only those complexes that contain completely complementary sequences.

Background signals can be reduced by the use of detergents such as SDS, Sarcosyl, or TRITON X-100

(Sigma-Aldrich, St. Louis MO) , and/or a blocking agent, such as salmon sperm DNA. Hybridization

25   methods are described in detail in Ausubel (supra, units 2.8-2.11, 3.18-3.19 and 4-6-4.9) and Sambrook

et al. (1989; Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY)

        NSEQ can be extended utilizing a partial nucleotide sequence and employing various PCR-

based methods known in the art to detect upstream sequences such as promoters and other regulatory

elements. (See, e.g., Dieffenbach and Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring

30   Harbor Press, Plainview NY). Additionally, one may use an XL-PCR kit (PE Biosystems, Foster City

CA), nested primers, and commercially available cDNA libraries (Life Technologies, Rockville MD) or

genomic libraries (Clontech, Palo Alto CA) to extend the sequence. For all PCR-based methods,

primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis

software (National Biosciences, Plymouth MN) or another program, to be about 18 to 30 nucleotides in

35   length, to have a GC content of about 50%, and to form a hybridization complex at temperatures of

about 68°C to 72°C.

In another aspect of the invention, NSEQ can be cloned in recombinant DNA molecules that direct the expression of PSEQ, or structural or functional portions thereof, in host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or

5 a functionally equivalent amino acid sequence may be produced and used to express the polypeptide encoded by NSEQ. The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter the nucleotide sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic

10 oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In order to express a biologically active polypeptide, NSEQ, or derivatives thereof, may be inserted into an expression vector, i.e., a vector which contains the elements for transcriptional and

15 translational control of the inserted coding sequence in a particular host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions. Methods which are well known to those skilled in the art may be used to construct such expression vectors. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, supra; and Ausubel,

20 supra).

A variety of expression vector/host cell systems may be utilized to express NSEQ. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with baculovirus vectors; plant cell systems transformed with viral

25 or bacterial expression vectors; or animal cell systems. For long term production of recombinant proteins in mammalian systems, stable expression in cell lines is preferred. For example, NSEQ can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable or visible marker gene on the same or on a separate vector. The invention is not to be limited by the vector or host cell employed.

30 In general, host cells that contain NSEQ and that express PSEQ may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or amino acid sequences. Immunological methods for detecting and

35 measuring the expression of PSEQ using either specific polyclonal or monoclonal antibodies are known

in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS).

Host cells transformed with NSEQ may be cultured under conditions for the expression and recovery of the polypeptide from cell culture. The polypeptide produced by a transgenic cell may be

5   secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing NSEQ may be designed to contain signal sequences which direct secretion of the polypeptide through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the

10 inserted sequences or to process the expressed polypeptide in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the polypeptide may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational

15 activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the ATCC (Manassas VA) and may be chosen to ensure the correct modification and processing of the expressed polypeptide.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences are ligated to a heterologous sequence resulting in translation of a fusion polypeptide containing heterologous polypeptide moieties in any of the aforementioned host systems. Such

20 heterologous polypeptide moieties facilitate purification of fusion polypeptides using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase, maltose binding protein, thioredoxin, calmodulin binding peptide, 6-His, FLAG, c-myc, hemaglutinin, and monoclonal antibody epitopes.

In another embodiment, the nucleic acid sequences are synthesized, in whole or in part, using

25 chemical or enzymatic methods well known in the art (Caruthers et al. (1980) Nucleic Acids Symp Ser (7) 215-233; Ausubel, supra). For example, peptide synthesis can be performed using various solid-phase techniques (Roberge et al. (1995) Science 269:202-204), and machines such as the ABI 431A Peptide synthesizer (PE Biosystems) can be used to automate synthesis. If desired, the amino acid sequence may be altered during synthesis and/or combined with sequences from other proteins to

30 produce a variant protein.

In another embodiment, the invention entails a substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:35 and fragments thereof.

SCREENING, DIAGNOSTICS AND THERAPEUTICS

The polynucleotide sequences can be used in diagnosis, prognosis, treatment, prevention, and

35 selection and evaluation of therapies for atherosclerosis including, but not limited to, stroke, myocardial

infarction, hypertension, transient cerebral ischemia, mesenteric ischemia, coronary artery disease, angina pectoris, peripheral vascular disease, intermittent claudication, renal artery stenosis, and hypertension.

The polynucleotide sequences may be used to screen a library of molecules for specific binding
5 affinity. The assay can be used to screen a library of DNA molecules, RNA molecules, PNAs, peptides, ribozymes, antibodies, agonists, antagonists, immunoglobulins, inhibitors, proteins including transcription factors, enhancers, repressors, and drugs and the like which regulate the activity of the polynucleotide sequence in the biological system. The assay involves providing a library of molecules, combining the polynucleotide sequence or a fragment thereof with the library of molecules under
10 conditions suitable to allow specific binding, and detecting specific binding to identify at least one molecule which specifically binds the polynucleotide sequence.

Similarly the polypeptide or a portion thereof may be used to screen libraries of molecules in any of a variety of screening assays. The portion of the polypeptide employed in such screening may be free in solution, affixed to an abiotic or biotic substrate (e.g. borne on a cell surface), or located
15 intracellularly. Specific binding between the polypeptide and molecule may be measured. The assay can be used to screen a library of DNA molecules, RNA molecules, PNAs, peptides, mimetics, ribozymes, antibodies, agonists, antagonists, immunoglobulins, inhibitors, peptides, polypeptides, drugs and the like, which specifically bind the polypeptide. One method for high throughput screening using very small assay volumes and very small amounts of test compound is described in Burbaum et
20 al. USPN 5,876,946, incorporated herein by reference, which screens large numbers of molecules for enzyme inhibition or receptor binding.

In one preferred embodiment, the polynucleotide sequences are used for diagnostic purposes to determine the absence, presence, and excess expression of the polypeptide. The polynucleotides may be at least 18 nucleotides long and consist of complementary RNA and DNA molecules, branched
25 nucleic acids, and/or peptide nucleic acids (PNAs). In one alternative, the polynucleotides are used to detect and quantify gene expression in samples in which expression of NSEQ is correlated with disease. In another alternative, NSEQ can be used to detect genetic polymorphisms associated with a disease. These polymorphisms may be detected in the transcript cDNA.

The specificity of the probe is determined by whether it is made from a unique region, a
30 regulatory region, or from a conserved motif. Both probe specificity and the stringency of diagnostic hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring, exactly complementary sequences, allelic variants, or related sequences. Probes designed to detect related sequences should preferably have at least 75% sequence identity to any of the polynucleotides encoding PSEQ.

35   Methods for producing hybridization probes include the cloning of nucleic acid sequences into

vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by adding RNA polymerases and labeled nucleotides. Hybridization probes may incorporate nucleotides labeled by a variety of reporter groups including, but not limited to, radionuclides such as ^{32}P or ^{35}S, enzymatic labels such as alkaline

5   phosphatase coupled to the probe via avidin/biotin coupling systems, fluorescent labels, and the like. The labeled polynucleotide sequences may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; and in microarrays utilizing samples from subjects to detect altered PSEQ expression.

         NSEQ can be labeled by standard methods and added to a sample from a subject under

10   conditions for the formation and detection of hybridization complexes. After incubation the sample is washed, and the signal associated with hybrid complex formation is quantitated and compared with a standard value. Standard values are derived from any control sample, typically one that is free of the suspect disease. If the amount of signal in the subject sample is altered in comparison to the standard value, then the presence of altered levels of expression in the sample indicates the presence of the

15   disease. Qualitative and quantitative methods for comparing the hybridization complexes formed in subject samples with previously established standards are well known in the art.

         Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual subject. Once the presence of disease is established and a treatment protocol is initiated, hybridization or

20   amplification assays can be repeated on a regular basis to determine if the level of expression in the subject begins to approximate that which is observed in a healthy subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to many years.

         The polynucleotides may be used for the diagnosis of a variety of diseases associated with

25   atherosclerosis. These include, but are not limited to, stroke, myocardial infarction, hypertension, transient cerebral ischemia, mesenteric ischemia, coronary artery disease, angina pectoris, peripheral vascular disease, intermittent claudication, renal artery stenosis, and hypertension.

         The polynucleotides may also be used as targets in a microarray. The microarray can be used to monitor the expression patterns of large numbers of genes simultaneously and to identify splice

30   variants, mutations, and polymorphisms. Information derived from analyses of the expression patterns may be used to determine gene function, to understand the genetic basis of a disease, to diagnose a disease, and to develop and monitor the activities of therapeutic agents used to treat a disease. Microarrays may also be used to detect genetic diversity, single nucleotide polymorphisms which may characterize a particular population, at the genome level.

35   In yet another alternative, polynucleotides may be used to generate hybridization probes useful

in mapping the naturally occurring genomic sequence. Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data as described in Heinz-Ulrich et al. (In: Meyers, supra, pp. 965-968).

In another embodiment, antibodies or Fabs comprising an antigen binding site that specifically

5 binds PSEQ may be used for the diagnosis of diseases characterized by the over-or-under expression of PSEQ. A variety of protocols for measuring PSEQ, including ELISAs, RIAs, and FACS, are well known in the art and provide a basis for diagnosing altered or abnormal levels of expression. Standard values for PSEQ expression are established by combining samples taken from healthy subjects, preferably human, with antibody to PSEQ under conditions for complex formation The amount of

10 complex formation may be quantitated by various methods, preferably by photometric means. Quantities of PSEQ expressed in disease samples are compared with standard values. Deviation between standard and subject values establishes the parameters for diagnosing or monitoring disease. Alternatively, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PSEQ specifically compete with a test compound for binding the polypeptide. Antibodies

15 can be used to detect the presence of any peptide which shares one or more antigenic determinants with PSEQ. In one aspect, the anti-PSEQ antibodies of the present invention can be used for treatment or monitoring therapeutic treatment for atherosclerosis.

In another aspect, the NSEQ, or its complement, may be used therapeutically for the purpose of expressing mRNA and polypeptide, or conversely to block transcription or translation of the mRNA.

20 Expression vectors may be constructed using elements from retroviruses, adenoviruses, herpes or vaccinia viruses, or bacterial plasmids, and the like. These vectors may be used for delivery of nucleotide sequences to a particular target organ, tissue, or cell population. Methods well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences or their complements. (See, e.g., Maulik et al. (1997) Molecular Biotechnology, Therapeutic Applications and

25 Strategies, Wiley-Liss, New York NY.) Alternatively, NSEQ, or its complement, may be used for somatic cell or stem cell gene therapy. Vectors may be introduced in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors are introduced into stem cells taken from the subject, and the resulting transgenic cells are clonally propagated for autologous transplant back into that same subject. Delivery of NSEQ by transfection, liposome injections, or polycationic amino polymers may be achieved using methods

30 which are well known in the art and described in Goldman et al. (1997; Nature Biotechnol 15:462-466). Additionally, endogenous NSEQ expression may be inactivated using homologous recombination methods which insert an inactive gene sequence into the coding region or other targeted region of NSEQ. (See, e.g. Thomas et al. (1987) Cell 51:503-512.)

Vectors containing NSEQ can be transformed into a cell or tissue to express a missing

35 polypeptide or to replace a nonfunctional polypeptide. Similarly a vector constructed to express the

complement of NSEQ can be transformed into a cell to downregulate the overexpression of PSEQ. Complementary or antisense sequences may consist of an oligonucleotide derived from the transcription initiation site; nucleotides between about positions -10 and +10 from the ATG are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful

5   because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee et al. In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco NY, pp. 163-177.)

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the cleavage of mRNA

10 and decrease the levels of particular mRNAs, such as those comprising the polynucleotide sequences of the invention. (See, e.g., Rossi (1994) Current Biology 4: 469-471.) Ribozymes may cleave mRNA at specific cleavage sites. Alternatively, ribozymes may cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The construction and production of ribozymes is well known in the art and is described in Meyers (supra).

15   RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiester linkages within the backbone of the molecule. Alternatively, nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine,

20 guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases, may be included.

Further, an antagonist, or an antibody that binds specifically to PSEQ may be administered to a subject to treat or prevent atherosclerosis. The antagonist, antibody, or fragment may be used directly to inhibit the activity of the polypeptide or indirectly to deliver a therapeutic agent to cells or tissues

25 which express the PSEQ. An immunoconjugate comprising a PSEQ binding site of the antibody or the antagonist and a therapeutic agent may be administered to a subject in need to treat or prevent disease. The therapeutic agent may be a cytotoxic agent selected from a group including, but not limited to, abrin, ricin, doxorubicin, daunorubicin, taxol, ethidium bromide, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicine, dihydroxy anthracin dione, actinomycin D, diphteria toxin,

30 Pseudomonas exotoxin A and 40, radioisotopes, and glucocorticoid.

Antibodies to PSEQ may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, such as those which inhibit dimer formation, are especially preferred for therapeutic use.

35 Monoclonal antibodies to PSEQ may be prepared using any technique which provides for the

production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma, the human B-cell hybridoma, and the EBV-hybridoma techniques. In addition, techniques developed for the production of chimeric antibodies can be used. (See, e.g., Pound (1998) Immunochemical Protocols, Methods Mol Biol Vol 80). Alternatively, techniques described for the
5 production of single chain antibodies may be employed. Fabs which contain specific binding sites for PSEQ may also be generated. Various immunoassays may be used to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art.

Yet further, an agonist of PSEQ may be administered to a subject to treat or prevent a disease
10 associated with decreased expression, longevity or activity of PSEQ.

An additional aspect of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic applications discussed above. Such pharmaceutical compositions may consist of PSEQ or antibodies, mimetics, agonists, antagonists, or inhibitors of the polypeptide. The compositions may be
15 administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a subject alone or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number
20 of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing
25 of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also
30 be used to determine the concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating and
35 contrasting the ED_{50} (the dose therapeutically effective in 50% of the population) and LD_{50} (the dose

lethal to 50% of the population) statistics. Any of the therapeutic compositions described above may be applied to any subject in need of such therapy, including, but not limited to, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

EXAMPLES

5   It is to be understood that this invention is not limited to the particular devices, machines, materials and methods described. Although particular embodiments are described, equivalent embodiments may be used to practice the invention. The described embodiments are provided to illustrate the invention and are not intended to limit the scope of the invention which is limited only by the appended claims.

10 I   cDNA Library Construction

The cDNA library SMCCNOS01 was selected as an example to demonstrate the construction of cDNA libraries from which the polynucleotides associated with known atherosclerosis-associated genes were derived. The SMCCNOS01 subtracted coronary artery smooth muscle cell library was constructed using 7.56×10^6 clones from the SMCCNOT02 library and was subjected to two rounds of
15 subtraction hybridization for 48 hours with 6.12×10^6 clones from SMCCNOT01. The SMCCNOT02 library was constructed using RNA isolated from coronary artery smooth muscle cells removed from a 3-year-old Caucasian male. The cells were treated for 20 hours with TNFα and IL-1β at 10ng/ml each. The SMCCNOT01 was constructed using RNA isolated from untreated coronary artery smooth muscle cells from the same donor. Subtractive hybridization conditions were based on the methodologies of
20 Swaroop et al. (1991; Nucleic Acids Res 19:1954) and Bonaldo et al. (1996; Genome Res 6:791).

For both cDNA libraries, SMCCNOT01 and SMCCNOT02, the frozen coronary artery smooth muscle cells (50-100 mg) were homogenized in GTC buffer (4.0M guanidine thiocyanate, 0.1M Tris-HCl pH 7.5, 1% 2-mercaptoethanol). Two volumes of binding buffer (0.4M LiCl, 0.1M Tris-HCl pH 7.5, 0.02M EDTA) were added, and the resulting mixture was vortexed at 13,000 rpm. The supernatant
25 was removed and combined with Oligo d(T)$_{25}$ bound streptavidin particles (MPG). After rotation at room temperature, the mRNA-Oligo d(T)$_{25}$ bound streptavidin particles were separated from the supernatant, washed twice with hybridization buffer I (0.15M NaCl, 0.01M Tris-HCl pH8.0, 1mM EDTA, 0.1% lauryl sarcosinate) using magnetic separation at each step to remove the supernatant from the particles. Bound mRNA was eluted from the particles with release solution and heated to 65°C.
30 The supernatant containing eluted mRNA was magnetically separated from the particles and used to construct the cDNA libraries.

The RNA was used for cDNA synthesis and construction of the cDNA library according to the recommended protocols in the SUPERSCRIPT plasmid system (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech (APB), Piscataway NJ),
35 and those cDNAs exceeding 400 bp were ligated into pINCY plasmid (Incyte Genomics, Palo Alto

CA). Recombinant plasmids were transformed into DH5α competent cells or ELECTROMAX cells (Life Technologies).

II   Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL PREP 96 plasmid kit
5 (Qiagen, Valencia CA). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile TERRIFIC BROTH media (BD Biosciences, Sparks MD) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cells were cultured for 19 hours and then lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml distilled water, and samples were transferred to a
10 96-well block for storage at 4° C.

The cDNAs were prepared using a MICROLAB 2200 System (Hamilton, Reno NV) in combination with the DNA ENGINE thermal cycler (MJ Research, Watertown MA). cDNAs were sequenced by the method of Sanger and Coulson (1975; J Mol Biol 94:441f) using ABI PRISM 377 (PE Biosystems) or MEGABACE 1000 sequencing systems (APB).
15   Most of the sequences disclosed herein were sequenced using standard ABI protocols and kits (PE Biosystems) at solution volumes of 0.25x -1.0x concentrations. In the alternative, some of the sequences disclosed herein were sequenced using solutions and dyes from APB.

III   Selection, Assembly, and Characterization of Sequences

The sequences used for co-expression analysis were assembled from EST sequences, 5' and 3'
20 longread sequences, and full length coding sequences. Selected assembled sequences were expressed in at least three cDNA libraries.

The assembly process is described as follows. EST sequence chromatograms were processed and verified. Quality scores were obtained using PHRED (Ewing et al. (1998) Genome Res 8:175-185; Ewing and Green (1998) Genome Res 8:186-194), and edited sequences were loaded into a relational
25 database management system (RDBMS). The sequences were clustered using BLAST with a product score of 50. All clusters of two or more sequences created a bin which represents one transcribed gene.

Assembly of the component sequences within each bin was performed using a modification of Phrap, a publicly available program for assembling DNA fragments (Green, P. University of Washington, Seattle WA). Bins that showed 82% identity from a local pair-wise alignment between
30 any of the consensus sequences were merged.

Bins were annotated by screening the consensus sequence in each bin against public databases, such as GBpri and GenPept from NCBI. The annotation process involved a FASTn screen against the GBpri database in GenBank. Those hits with a percent identity of greater than or equal to 75% and an alignment length of greater than or equal to 100 base pairs were recorded as homolog hits. The residual
35 unannotated sequences were screened by FASTx against GenPept. Those hits with an E value of less

than or equal to 10^{-8} were recorded as homolog hits.

Sequences were then reclustered using BLASTn and Cross-Match, a program for rapid amino acid and nucleic acid sequence comparison and database search (Green, supra), sequentially. Any BLAST alignment between a sequence and a consensus sequence with a score greater than 150 was
5   realigned using cross-match. The sequence was added to the bin whose consensus sequence gave the highest Smith-Waterman score (Smith et al. (1992) Incyte Genomics 5:35-51) amongst local alignments with at least 82% identity. Non-matching sequences were moved into new bins, and assembly processes were repeated.

IV   Coexpression Analyses of Atherosclerosis-Associated Genes

10   Sixty-six known atherosclerosis-associated genes were selected to identify novel genes that are closely associated with atherosclerosis. The known atherosclerosis-associated genes which were examined in this analysis and brief descriptions of their functions are listed in Table 3.

Table 3. Descriptions of Known Atherosclerosis-Associated Genes

GENE	DESCRIPTION AND REFERENCES
Human 22kDa smooth muscle protein (SM22)	Smooth muscle cell-specific gene which is down-regulated during smooth muscle cell dedifferentiation as part of atherogenic process (Sobue et al. (1998) Horm Res 50:15-24; Sobue et al. (1999) Mol Cell Biochem 190:105-18)
calponin (CNN1)	Calponin is smooth muscle-specific and may mediate smooth muscle contractility through it's binding of the amino-terminal end of the myosin regulatory light chain. Involved in phenotypic modulation of smooth muscle cells, a feature of atherosclerosis (Szymanski et al. (1999) Biochemistry 38:3778-84)
desmin (DES)	Contractile component of myofibrils in differentiated smooth muscle cells. Regarded as a marker for smooth muscle cells (Shi et al. (1997) Circulation 95:2684-93)
smooth muscle myosin heavy chain (MYH11)	Contractile component of myofibrils in differentiated smooth muscle cells. Regarded as a marker for smooth muscle cells (Sobue et al. (1999) Mol Cell Biochem 190:105-18)
alpha tropomyosin (TPM1)	Contractile component of myofibrils in differentiated smooth muscle cells (Sobue et al. (1999) Mol Cell Biochem 190:105-18; Kashiwada et al. (1997) J Biol Chem 272:15396-404)
Human tissue inhibitor of metalloprotein-ase 3 (TIMP3)	TIMPs control the activity of matrix metalloproteinases and are important in local matrix remodeling of vasculature. Atheroma extracts shown to have 5x higher TIMP3 expression levels than non-atherosclerotic tissue. Abundant TIMP1,2, 3 expression noted in plaque macrophages and smooth muscle cells. PDGF and TGFbeta augment TIMP3 expression. TIMP3 possible important role in plaque stability (Fabunmi et al. (1998) Circ Res 83:270-8)

	Human tissue inhibitor of metalloprotein-ase-2 (TIMP-2)	TIMPs control the activity of matrix metalloproteinases and are important in local matrix remodeling of vasculature. Abundant TIMP1,2, 3 expression noted in plaque macrophages and smooth muscle cells. Expression of TIMP2 is greatly increased during neointima formation in organ cultures of human saphenous vein (Kranzhofer et al. (1999) Arterioscler Thromb Vasc Biol 19:255-65)
5	Human tissue inhibitor of metalloprotein-ase-4 (TIMP4)	TIMPs control the activity of matrix metalloproteinases and are important in local matrix remodeling of vasculature (Greene et al. (1996) J Biol Chem 271:30375-80)
10	pro alpha 1(I) collagen (COL1A1)	Member of family of fibrous structural proteins. Most abundant structural component of the extracellular matrix. Secreted as procollagen and converted to collagen by matrix metalloproteinases. Collagens are important in atherosclerosis for promoting platelet aggregation and for providing sites for platelet adhesion to the vessel wall (Wen et al. (1999) Arterioscler Thromb Vasc Biol 19:519-24)
	collagen alpha-2 type I (COL1A2)	see COL1A1 above
	COL6A1	see COL1A1 above
15	procollagen alpha 2(V) (COL5A2)	see COL1A1 above
	collagen VI alpha-2 (COL6A2)	see COL1A1 above
20	type VI collagen alpha3 (COL6A3)	see COL1A1 above
	pro-alpha-1 type 3 collagen (COL3A1)	see COL1A1 above
25	pro-alpha-1 (V) collagen (COL3A1	see COL1A1 above
30	collagenase type IV/ matrix metalloprotein-ase 9/gelatinase B (MMP9)	Contributes to the degradation of vascular wall/smooth muscle cells associated with local matrix remodeling. Expression of metalloproteinases controlled by tissue inhibitors of metalloproteinases (TIMPs). Balance between MMP and TIMP expression becomes distorted during onset and progression of atherosclerosis. MMP9 localized to lesional macrophages, along with MMP-1, MMP-2, MMP-3. Rabbit aortic macrophage foam cells express immunoreactive MMP-9 (Moreau et al. (1999) Circulation 99:420-426; Zaltsman et al. (1997) Atherosclerosis 130:61-70)
	matrix Gla protein (MGP)	Role in active calcification of vascular smooth muscle cells, suggested by expression study on VSMC in vitro differentiation study. Calcifying phenotype associated with high MGP levels. MGP knockout mice develop to term, but die up to 2 months after birth due to extensive calcification of the arteries, causing blood vessel rupture (Luo et al. (1997) Nature 386:78-81; Mori et al. (1998) FEBS Lett 433:19-22)

	cathepsin K (CTSK)	Nonmetalloenzyme, potent elastase present in advanced atherosclerotic plaques. Contributes to the breakdown of components of vascular extracellular matrix, reducing tensile strength, increasing plaque vulnerability (Sukhova et al. (1998) J Clin Invest 102:576-83)
	fibrinogen beta chain gene (FGB)	Component of fibrin in the extracellular matrix. Fibrin deposition is an integral part of advanced atherosclerotic lesion development. Variation at the beta fibrinogen locus associated with peripheral atherosclerosis (Sueishi et al. (1998) Semin Thromb Hemost 24:255-260; Fowkes et al. (1992) Lancet 339:693-696)
5	fibrinogen beta chain gene (FGG)	Participant in adhesion and aggregation of platelets which occurs through binding of platelet receptors. FGG carries the main binding site for the platelet receptor binding. Mutations in FGG associated with clotting defects and thrombotic tendency. Fibrin deposition is an integral part of advanced atherosclerotic lesion development (Sueishi et al. (1998) Semin Thromb Hemost 24:255-60; Cote et al. (1998) Blood 92:2195-2212)
	pre-pro-von Willebrand factor (VWF)	Blood glycoprotein involved in normal hemostasis. Mediates adhesion of platelets to sites of vascular damage. Also acts as a cofactor in factor VIII activity in blood coagulation. Increased levels of VWF are found in atherosclerosis and in several of its major risk factors, including hypercholesterolemia, diabetes, obesity, hypertension. Levels serve as a predictor of adverse clinical outcome following vascular surgery, possibly as an indicator of thrombus formation (Sadler (1998) Annu Rev Biochem 67:395-424; Blann et al. (1994) Eur J Vasc Surg 8:10-15; Kessler et al. (1998) Diabetes Metab 24:327-36; Folsom et al. (1997) Circulation 96:1102-1108)
10	coagulation factor II/ prothrombin (F2)	Central role in blood hemostasis by regulating platelet aggregation and blood coagulation. Converts fibrinogen to fibrin in the final stage of clotting cascade. Promotes cellular chemotaxis and proliferation, extracellular matrix turnover and release of inflammatory cytokines (Goldsack et al. (1998) Int J Biochem Cell Biol 30:641-646)
	coagulation factor XII (F12)	Activation of blood coagulation is an important part of post-vascular injury with initiation of atherosclerotic lesion formation and contributes to thrombosis in advanced stage atherosclerosis (Sueishi et al. (1998) Semin Thromb Hemost 24:255-260)
15	coagulation factor VII (F7)	Central role in coagulation, influences plasma triglyceride levels, a risk factor in atherosclerosis. Epidemiological studies have linked F7 with cardiovascular risk/ atherothrombotic tendency (Ghaddar et al. (1998) Circulation 98:2815-2821; Koenig (1998) Eur Heart J 19:C39-43; Folsom et al. (1997) Circulation 96:1102-1108)
20	platelet endothelial cell adhesion molecule (PECAM-1)	Signalling molecule in the migration of cells as part of the pathophysiology of vascular occulsive diseases such as atherosclerosis. Analysis of endothelial/monocyte co-cultures indicates oxidative stress induces transendothelial migration of monocytes as a result of phosphorylation of PECAM-1 (Rattan et al. (1997) Am J Physiol 273:E453-61)

	lipoprotein-associated coagulation inhibitor (LACI)	Natural anticoagulant, inhibits factor VII/tissue factor complexes. Role in regulating coagulation in atherosclerotic plaques. Circulates in association with plasma lipoproteins VLDL, HDL and LDL. In situ expression studies indicate TFPI is expressed in adventitial layer of large arteries, and in atherosclerotic vessels is expressed by macrophages in focal areas throughout the plaque (Drew et al. (1997) Lab Invest 77:291-298; Sandset (1996) Haemostasis 26:154-165)
5	antithrombin III variant (AT3)	ATIII is the sole blood component through which heparin exerts its anti-coagulation effect. Deficiency in ATIII causes recurrent venous thrombosis and pulmonary embolism and can be inherited in autosomal dominant fashion (Hultin et al. (1988) Thromb Haemost 59:468-73; Lane et al. (1996) Blood Rev 10:59-74)
10	plasminogen activator inhibitor-1 (PAI-1)	Major physiological inhibitor of fibrinolysis. Plasma levels correlate with incidence of MI and venous thrombosis. Both adipocytes and endothelial cells produced PAI, possibly under the control of PPARG, as demonstrated using recombinant PPARG expression constructs in endothelial cell lines. Increased expression of PAI observed in coronary heart disease. 4G polymorphism in promotor causes increased PAI expression associated with MI in some studies (Eriksson et al. (1995) Proc Natl Acad Sci 92:1851-5; Marx et al. (1999) Arterioscler Thromb Vasc Biol 19:546-551)
	lipoprotein lipase (LPL)	Hydrolises triglyceride in chylomicrons and therefore regulates metabolism of circulating lipoproteins. Appears to have an atherogenic effect on the arterial wall due to its ability to alter the properties of LDL. Increased activity of LPL is found in atherosclerotic arteries when compared to normal. Expressed by macrophages in atherosclerotic lesions. Mutations in LPL responsible for familial hyperchol-esterolemia and premature atherosclerosis (Fisher et al. (1997) Atherosclerosis 135:145-159; Goldberg (1996) J Lipid Res 37:693-707; Gerdes et al. (1997) Circulation 96:733-740)
15	alpha-2-macroglobulin (A2M)	Foam cell formation - retains LDL cholesterol in the lipid core of atherosclerotic plaque (Llorente et al. (1998). Rev Esp Cardiol 51:633-641)
	apolipoprotein AI (APOA1)	Participates in reverse cholesterol transport from tissues to the liver. Promotes cholesterol efflux from tissues and acts as a cofactor for lecithin cholesterol acyltransferase (LCAT). Mutations in ApoA1 and of ApoAI/CIII/AIV gene cluster assoc with atherosclerosis. Transgenic mice expressing high plasma APOAI levels are protected from fatty streak development with a high atherogenic diet (Gordon et al. (1989) Circulation 79:8-15; Rubin et al. (1991) Nature 353:265-7; Karathanasis et al. (1987) Proc Natl Acad Sci 84:7198-7202)
	apolipoprotein AII (APOA2)	Major component of HDL. Appears to have an opposite effect to that of APOAI, though exact function unknown. APOAII may have ability to convert HDL from an anti- to a pro-inflammatory particle, with paraoxonase having a role in this transformation process. Plasma APOAII levels significantly associated with plasma free fatty acid levels. Transgenic mice expressing varying levels of APOAII show increased atherosclerotic lesions than wt when fed an atherogenic diet. Possible interaction between diet/genotype and atherogenic potential (Escola-Gil et al. (1998) J Lipid Res 39:457-462; Warden et al. (1993) Proc Natl Acad Sci 90:10886-10890)

	apolipoprotein B-100 (APOB)	Main apolipoprotein of chylomicrons and low density lipoproteins. Mutations in APOB100 underly familial defective apolipoprotein B-100 in which patients suffer from premature atherosclerosis. Mutations result in defect in binding of LDL to LDL receptor, and accumulation of plasma LDL. High-expressing APOB transgenic mice exibit elevated VLDL-LDL cholesterol and atherogenic lesions (Callow et al. (1995) J Clin Invest 96:1639-1646; Brasaemle et al. (1997) J Biol Chem 272:9378-9387)
	lipoprotein apoCII (APOC2)	Role in lipoprotein metabolism. Cofactor in the activity of lipoprotein lipase the enzyme that hydrolyzes triglycerides in plasma and transfers the fatty acids to tissues. Mutations in APOC2 responsible for hyperlipoproteinemia 1B, similar to lipoprotein lipase deficiency (Cox et al. (1978) N Engl J Med 299:1421-1424; Arimoto et al. (1998) J Lipid Res 39:143-151)
5	pre-apolipoprotein CIII (APOC3)	Inhibits lipoprotein lipase and hepatic lipase, decreases uptake of lymph chylomicrons by hepatic cells. APOA3 possibly delays breakdown of triglyceride rich particles. SstI RFLP in apoCIII is associated with plasma triglyceride and apoCIII levels and hyperlipidemic phenotypes (Henderson et al. (1987) Hum Genet 75:62-65)
	apolipoprotein apoC-IV (APOC4)	APOC4 is a lipid-binding protein that has the potential to alter lipid metabolism. Human APOC4 transgenic mice are hypertriglyceridaemic compared to normal controls (Allan et al. (1996) J Lipid Res 37:1510-1518)
10	macrophage scavenger receptor type I (MSR1)	Mediates binding, internalisation and processing of negatively-charged macromolecules. Implicated in the pathological deposition of cholesterol in arterial walls during atherogenesis (Han et al. (1998) Hum Mol Genet 7:1039-1046)
15	Human antigen CD36 gene (CD36)	Acts as a scavenger receptor for oxidised LDL. Transient regulation under control of M-CSF during monocyte-macrophage differentiation increases foam cell accumulation, Possible role in atherogenesis: increased M-CSF levels detected in atherosclerotic lesions in rabbits and humans. (Huh et al. (1996) Blood 87:2020-2028; Aitman et al. (1999) Nat Genet 21:76-83)
	serum amyloid P component (SAP)	Plasma glycoprotein expressed in atherosclerotic lesions. Interacts with lipoproteins in specific manner (Li et al. (1995) Arterioscler Thromb Vasc Biol 15:252-257; Li et al. (1998) Biochem Biophys Res Commun 244:249-252)
	carboxyl ester lipase gene (CEL)	CEL gene expression increases in presence of oxidised and native LDL in vitro. It is expressed in the vessel wall and in aortic extracts - may interact with cholesterol to modulate progression of atherosclerosis (Li et al. (1998) Biochem J 329:675-679)
20	paraoxonase 1 (PON1)	Serum esterase exclusively associated with high-density lipoproteins; it might confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids in oxidized low-density lipoproteins. PON1 gln192-to-arg polymorphism associated with CAD. Association between PON1 genetic variation and plasma LDL, HDL and non-HDL and apoB levels in genetically isolated Alberta Hutterite population. When fed on a high-fat, high-cholesterol diet, PON1-null mice were more susceptible to atherosclerosis than wild-type (Serrato et al. (1995) J Clin Invest 96:3005-3008; Boright et al. (1998) Atherosclerosis 139:131-136; Shih et al. (1998) Nature 394:284-287)

	paraoxonase 2 (PON2)	Serum esterase exclusively associated with high-density lipoproteins; it might confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids in oxidized low-density lipoproteins. Common polymorphism at codon 311 (cys-ser) in PON2 associated with CHD alone and synergistically with the 192 polymorphism in PON1 in Asian Indians. Association between genetic variation in PON2 and plasma cholesterol and apolipoprotein A1 in genetically isolated Alberta Hutterite population (Sanghera et al. (1998) Am J Hum Genet 62:36-44; Boright et al. (1998) Atherosclerosis 139:131-136)
	paraoxonase 3 (PON3)	Serum esterase exclusively associated with high-density lipoproteins; it might confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids in oxidized low-density lipoproteins. Other members PON2, 3 associated with CHD and cholesterol levels (Laplaud et al. (1998) Clin Chem Lab Med 36:431-441)
5	LDL-receptor related protein (LRP1)	Possible important role in atherosclerotic lesion development. Abundant expression of mRNA and protein found in vascular smooth muscle cells and macrophages of early and advanced atherosclerotic lesions. Receptor for uptake of ApoE-containing lipoprotein particles (Beisiegel et al. (1989) Nature 341:162-164; Hiltunen et al. (1998) Atherosclerosis 137:S81-88)
10	hepatic triglyceride lipase (HTGL)	Hepatic lipase is involved in cholesterol efflux. Downstream of cholesterol ester transfer protein in pathway: acts on triglyceride-rich HDL to promote formation of smaller HDL particles - effectors of cellular cholesterol efflux (Fan et al.(1998) J Atheroscler Thromb 5:41-45; Santamarina-Fojo et al. (1998) Curr Opin Lipidol 9:211-219)
15	3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGCR)	Catalyses rate limiting step in cholesterol biosynthesis as well as being involved in other systems (eg. primordial germ cell migration). Expression of HMG CoA reductase is regulated by oxysterols via sterol-regulatory element in the promotor, as is found in APOE. Target for cholesterol-lowering therapies: prevastatin, "statins" (Bocan et al. (1998) Atherosclerosis 139:21-30; Farnier et al. (1998) Am J Cardiol 82:3J-10J)
	very low density lipoprotein receptor (VLDLR)	Role in triglyceride metabolism. Marked induction of VLDLR expression observed in fatty streaks and plaques in rabbit atherosclerosis models (Hiltunen et al. (1998) Circulation 97:1079-1086)
20	Microsomal triglyceride transfer protein (MTP)	Catalyses transport of triglyceride, cholesterol ester and phospholipid between phospholipid surfaces. Mutations cause abetalipoproteinemia. Linkage found between MTP genotype and plasma triglyceride levels in a quantitative sib-pair analysis of female dizygotic twins. Inhibitors of MTP normalise atherogenic lipoprotein profiles in an atherosclerotic rabbit model (Wetterau et al. (1992) Science 258:999-1001; Austin et al. (1998) Am J Hum Genet 62:406-419; Wetterau et al. (1998) Science 282:751-754)
	perilipin (PLIN)	Lipid storage droplets of steroidogenic cells are surrounded by perilipins, family of phosphorylated proteins encoded by a single gene, detected in adipocytes and steroidogenic cells. Possible role in lipid metabolism (Brasaemle et al. (1997) J Biol Chem 272:9378-9387)

	endothelin-1 (EDN1)	Secretion of EDN1 coincides with the location of native and oxidised low density lipoproteins and occurs in a specific fashion suggesting that EDN1 may be involved in pathophysiological processes such as atherogenesis. Quantitative and qualitative immunohistochemical analysis of anti EDN1 antibodies in the wall layers of human arteries ex vivo suggest that EDN1 is normally expressed exclusively in endothelial cells. However, in cases of coronary artery disease and atherosclerosis, EDN1 expression is enhanced and can be found in the tunica media and vascular smooth muscle cells. Analysis of recombinant EDN1 expression in vitro suggests it influences vascular smooth muscle cell proliferation. Potent vasoconstriction properties (Unoki et al. (1999) Cell Tissue Res 295:89-99; Rossi et al. (1999) Circulation 99:1147-1155; Yoshizumi et al. (1998) Br J Pharmacol 125:1019-1027; Alberts et al. (1994) J Biol Chem 269: 10112-10118)
5	endothelin receptor A (EDNRA)	Mediates action of endothelin1 on vascular smooth muscle migration, proliferation and monocyte/endothelial cell interaction during initiation and progression of atherosclerotic lesion development (Kohno et al. (1998) J Cardiovasc Pharmacol 31:S84-9; Alberts et al. (1994) J Biol Chem 269:10112-10118)
	interleukin 6 (IL6)	Inflammatory cytokine present in arterial atherosclerotic wall which is upregulated by platelets to stimulate smooth muscle cell growth. Increased expression of IL6 in atherosclerotic aortas of APOE knockout vs aortas from aged-matched controls. Secretion levels of IL6 is positively associated with increased lesion surface area in APOE aortic tissue samples (Sukovich et al. (1998) Arterioscler Thromb Vasc Biol 18:1498-1505; Loppnow et al. (1998) Blood 91:134-141)
	interleukin 1 (IL1)	May contribute to regulation of local pathogenesis in the vessel wall by activation of the cytokine regulatory network. IL-1 antagonist inhibits platelet-induced cytokine production of smooth muscle cells (Loppnow et al. (1998) Blood 91: 134-141)
10	complement protein C8 alpha (C8A)	Complement activation of C8 shown to be an initial event in atherogenesis (Torzewski et al. (1996) Arterioscler Thromb Vasc Biol 16:673-677)
	complement component C9 (C9)	Complement activation of C9 shown to be an initial event in atherogenesis (Torzewski et al. (1996) Arterioscler Thromb Vasc Biol 16:673-677)
15	Prostaglandin D2 synthase (PTGDS)	Catalyses conversion of PGH2 to PGD2, a prostaglandin important in smooth muscle contraction/relaxation and potent inhibitor of platelet aggregation. Northern analysis shows strong specific expression in heart. Immunocytochemical localisation to myocardial and atrio endocardial cells, and accumulates in end-stage atherosclerotic plaques. High plasma levels detected in severe angina patients (Eguchi et al. (1997) Proc Natl Acad Sci 94:14689-14694)
	Annexin II/lipocortinII (ANX2)	Inhibits phospholipase A2 activity and hence the production of arachidonic acid, the precursor of the inflammatory mediators prostaglandins and leukotrienes. ANX2 is an important anti-inflammatory molecule. Independently binds plasminogen and t-PA and therefore suspected of having a role in atherogenesis. Binding of plasminogen to ANX2 is specifically inhibited by the excess atherogenic Lp(a) (Hajjar et al.(1998) J Investig Med 46(8): 364-369)

Annexin I/lipocortin (ANX1)	Inhibits phospholipase A2 activity and hence the production of arachidonic acid, the precursor of the inflammatory mediators prostaglandins and leukotrienes. ANXI is an important anti-inflammatory molecule (Wallner et al. (1986) Nature 320:77-81)
Prostaglandin-endoperoxide Synthase 2 (PTGS2)	Major mechanism for the regulation of prostaglandin synthesis. Arachidonic acid pathway. Role in inflammation and endothelial cell migration/angiogenesis. Regulated enzyme - major mediator of inflammation. Antiinflammatory glucocorticoids are potent inhibitors of this cyclooxygenase. Over expression of PTGS2 in vitro in rabbit epithelial cells causes increased adhesion to extracellular matrix proteins and inhibition of apoptosis, hallmarks of atherosclerotic plaque formation (Morham et al. (1995) Cell 83:473-482; O'Banion et al. (1992) Proc Natl Acad Sci 89:4888-4892; Tsujii et al. (1995) Cell 83:493-501)
insulin-like growth factor binding protein-1 (IGFBP-1)	A study of 218 individuals indicates free IGFBP1 levels are associated with high HDL cholesterol and more favourable cardiovascular outcome. The IGF1/IGFBP1 system found to be associated with cardiovascular risk and atherosclerosis (Janssen et al. (1998) Arterioscler Thromb Vasc Biol 18:277-282)
Secreted protein, acidic and rich in cysteine (SPARC)	Extracellular glycoprotein secreted by endothelial cells which has a suspected role in calcification of atherosclerotic plaques. Interacts with PDGF-B containing dimers and inhibits binding to its receptors. Expression of SPARC and PDGF is minimal in most adult tissues, but is enhanced following injury and advanced atherosclerotic lesions. Selective expression of SPARC causes rounding of adherent endothelial cells and influences extravasation of macromolecules (Raines et al. (1992) Proc Natl Acad Sci 89:1281-1285; Goldblum et al. (1994) Proc Natl Acad Sci 91:3448-3452)
Human NF-kappa-B transcription factor (NFkB)	Activated NF kappa B occurs in atherosclerotic lesions, and regulates the expression of gene important in recruitment of monocytes and inflammatory response. Responsible for cytokine production by smooth muscle cells during atherogenesis (Navab et al. (1995) Am J Cardiol 76:18C-23C; Hernandez-Presa et al. (1998) Am J Pathol 153:1825-1837; Thurberg et al. (1998) Curr Opin Lipidol 9:387-396; Brand et al. (1997) Arterioscler Thromb Vasc Biol 17:1901-1909)
angiotensinogen (AGT)	Concentration of angiotensinogen influences the renin-angiotensin system(RAS). Hypertensive mice carrying renin and angiotensinogen transgenes found to have higher total cholesterol levels on an atherogenic diet than their wt counterparts, and atherogenic lesions were 4x larger in surface area. Suggests hypertension induced by activated RAS is important atherogenic factor (Sugiyama et al. (1997) Lab Invest 76:835-842)
Nitric Oxide Synthase 3 (NOS3)	Mediates basal vasodilation. Regulates the production of nitric oxide, an important signal transduction component and scavenger of reactive oxygen species. Activity of NOS3 appears to be a factor in endothelin/endothelin receptor B mediated endothelial cell migration and angiogenesis. Polymorphism associated with smoking dependent coronary artery disease (Goligorsky et al. (1999) Clin Exp Pharmacol Physiol 26:269-271; Stroes et al. (1998) J Cardiovasc Pharmacol 32:S14-21; Sobue et al. (1998) Horm Res 50:15-24)

Nitric Oxide Synthase 2 (NOS2)	Mediates basal vasodilation. Regulates the production of nitric oxide, an important signal transduction component and scavenger of reactive oxygen species. NOS2, known as inducible NOS is expressed in most cells only after induction by immunologic and inflammatory stimuli, and is upregulated in pathological conditions such as atherosclerosis (Dusting et al. (1998) Clin Expt Pharmacol Phisiol 25:S34-41)

5   From a total of 45,233 assembled gene sequences, 34 novel genes were identified, SEQ ID NOs:1-34, that show strong association with 66 known atherosclerosis-associated genes. Initially, the degree of association was measured by probability values using a cutoff p value less than 0.00001. The sequences were further examined to ensure that the genes that passed the probability test had strong association with known atherosclerosis-associated genes. Details of the co-expression patterns for the

10   66 known and 34 novel atherosclerosis-associated polynucleotides are presented in Table 4. The entries in Table 4 are the negative log of the p-value (-log p) for the coexpression of the two genes. The novel atherosclerosis-associated polynucleotides identified are listed in the table by their SEQ ID NOs numbers, and the known genes, by their names or the abbreviations shown in Table 3.

V   Novel Genes Associated with Atherosclerosis

15   Using the co-expression analysis method, 34 novel atherosclerosis-associated polynucleotides were identified, SEQ ID NOs:1-34, that exhibit strong association, or co-expression, with 66 known atherosclerosis-associated genes.

Polynucleotides comprising the consensus sequences of SEQ ID NO:1-34 of the present invention were first identified from Incyte bins and assembled as described in Example III. BLAST

20   and other motif searches were performed for SEQ ID NOs:1-34 according to Example VI. The full length and 5'-complete sequences were translated and sequence identity was sought with known sequences.

SEQ ID NO:35 of the present invention was encoded by the nucleic acids of SEQ ID NO:11. SEQ ID NO:35 has 366 amino acids which are encoded by SEQ ID NO:11. Motif analyses of SEQ ID

25   NO:35 shows one potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at residue S343, two potential casein kinase II phosphorylation sites at residues S179 and T351, and four potential protein kinase C phosphorylation sites at residues T29, S85, T269, and T324. Additionally, SEQ ID NO:35 contains a potential sugar transport protein signature sequence from residues L201 to S217.

30 **VI   Homology Searching for Atherosclerosis-Associated Polynucleotides and Polypeptides**

The polynucleotide sequences, SEQ ID NO:1-34, and polypeptide sequence, SEQ ID NO:35, were queried against databases derived from sources such as GenBank and SwissProt. These databases, which contain previously identified and annotated sequences, were searched for regions of similarity using BLAST (Altschul, supra). BLAST searched for matches and reported only those that satisfied the

probability thresholds of 10^{-25} or less for nucleotide sequences and 10^{-8} or less for polypeptide sequences.

The polypeptide sequence was also analyzed for known motif patterns using MOTIFS, SPSCAN, BLIMPS, and HMM-based protocols. MOTIFS (Genetics Computer Group, Madison WI)

5   searches polypeptide sequences for patterns that match those defined in the Prosite Dictionary of Protein Sites and Patterns (Bairoch, supra) and displays the patterns found and their corresponding literature abstracts. SPSCAN (Genetics Computer Group) searches for potential signal peptide sequences using a weighted matrix method (Nielsen et al. (1997) Prot Eng 10:1-6). Hits with a score of 5 or greater were considered. BLIMPS uses a weighted matrix analysis algorithm to search for

10 sequence similarity between the polypeptide sequences and those contained in BLOCKS, a database consisting of short amino acid segments, or blocks of 3-60 amino acids in length, compiled from the PROSITE database (Henikoff; supra; Bairoch, supra), and those in PRINTS, a protein fingerprint database based on non-redundant sequences obtained from sources such as SwissProt, GenBank, PIR, and NRL-3D (Attwood et al. (1997) J Chem Inf Comput Sci 37:417-424). For the purposes of the

15 present invention, the BLIMPS searches reported matches with a cutoff score of 1000 or greater and a cutoff probability value of 1.0×10^{-3}. HMM-based protocols were based on a probabilistic approach and searched for consensus primary structures of gene families in the protein sequences (Eddy, supra; Sonnhammer, supra). More than 500 known protein families with cutoff scores ranging from 10 to 50 bits were selected for use in this invention.

20 **VII   Labeling of Probes and Hybridization Analyses**

Substrate Preparation

Nucleic acids are isolated from a biological source and applied to a substrate for standard hybridization protocols by one of the following methods. A mixture of target nucleic acids, a restriction digest of genomic DNA, is fractionated by electrophoresis through an 0.7% agarose gel in

25 1xTAE [Tris-acetate-ethylenediamine tetraacetic acid (EDTA)] running buffer and transferred to a nylon membrane by capillary transfer using 20x saline sodium citrate (SSC). Alternatively, the target nucleic acids are individually ligated to a vector and inserted into bacterial host cells to form a library. Target nucleic acids are arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane.

30 The membrane is placed on bacterial growth medium, LB agar containing carbenicillin, and incubated at 37°C for 16 hours. Bacterial colonies are denatured, neutralized, and digested with proteinase K. Nylon membranes are exposed to UV irradiation in a STRATALINKER UV-crosslinker (Stratagene) to cross-link DNA to the membrane.

In the second method, target nucleic acids are amplified from bacterial vectors by thirty cycles

35 of PCR using primers complementary to vector sequences flanking the insert. Amplified target nucleic

acids are purified using SEPHACRYL-400 beads (APB). Purified target nucleic acids are robotically arrayed onto a glass microscope slide (Corning Science Products, Corning NY). The slide is previously coated with 0.05% aminopropyl silane (Sigma-Aldrich) and cured at 110°C. The arrayed glass slide (microarray) is exposed to UV irradiation in a STRATALINKER UV-crosslinker (Stratagene).

5 Probe Preparation

cDNA probes are made from mRNA templates. Five micrograms of mRNA is mixed with 1 µg random primer (Life Technologies), incubated at 70°C for 10 minutes, and lyophilized. The lyophilized sample is resuspended in 50 µl of 1x first strand buffer (cDNA Synthesis systems; Life Technologies) containing a dNTP mix, [α-^{32}P]dCTP, dithiothreitol, and MMLV reverse transcriptase

10 (Stratagene), and incubated at 42°C for 1-2 hours. After incubation, the probe is diluted with 42 µl dH$_2$O, heated to 95°C for 3 minutes, and cooled on ice. mRNA in the probe is removed by alkaline degradation. The probe is neutralized, and degraded mRNA and unincorporated nucleotides are removed using a PROBEQUANT G-50 microcolumn (APB). Probes can be labeled with fluorescent markers, Cy3-dCTP or Cy5-dCTP (APB), in place of the radionucleotide, [^{32}P]dCTP.

15 Hybridization

Hybridization is carried out at 65°C in a hybridization buffer containing 0.5 M sodium phosphate (pH 7.2), 7% SDS, and 1 mM EDTA. After the substrate is incubated in hybridization buffer at 65°C for at least 2 hours, the buffer is replaced with 10 ml of fresh buffer containing the probes. After incubation at 65°C for 18 hours, the hybridization buffer is removed, and the substrate is

20 washed sequentially under increasingly stringent conditions, up to 40 mM sodium phosphate, 1% SDS, 1 mM EDTA at 65°C. To detect signal produced by a radiolabeled probe hybridized on a membrane, the substrate is exposed to a PHOSPHORIMAGER cassette (APB), and the image is analyzed using IMAGEQUANT data analysis software (APB). To detect signals produced by a fluorescent probe hybridized on a microarray, the substrate is examined by confocal laser microscopy, and images are

25 collected and analyzed using GEMTOOLS gene expression analysis software (Incyte Genomics).

VIII   Complementary Polynucleotides

Molecules complementary to the polynucleotide, or a fragment thereof, are used to detect, decrease, or inhibit gene expression. Although use of oligonucleotides comprising from about 18 to about 60 base pairs is described, the same procedure is used with larger or smaller fragments or their

30 derivatives (PNAs). Oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and SEQ ID NO:1-34 or fragments thereof To inhibit transcription by preventing promoter binding, a complementary oligonucleotide is designed to bind to the most unique 5' sequence, most preferably about 10 nucleotides before the initiation codon of the open reading frame. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the mRNA encoding the

35 polypeptide.

IX   Producti n of Specific Antibodies

The polypeptides encoded by SEQ ID NO:1-34, or portions thereof, substantially purified using polyacrylamide gel electrophoresis or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols as described in Pound (supra).

5   Alternatively, the amino acid sequence is analyzed using LASERGENE software (DNASTAR, Madison WI) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A Peptide

10   synthesizer (PE Biosystems) using fmoc-chemistry and coupled to keyhole limpet hemocyanin (KLH, Sigma-Aldrich) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (Ausubel, supra) to increase immunogenicity. Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with

15   radio-iodinated goat anti-rabbit IgG.

X   Screening Molecules for Specific Binding with the Polynucleotide or Polypeptide

The polynucleotide, or fragments thereof, or the polypeptide, or portions thereof, are labeled with ^{32}P-dCTP, Cy3-dCTP, or Cy5-dCTP (APB), or with BIODIPY or FITC (Molecular Probes, Eugene OR), respectively. Libraries of candidate molecules previously arranged on a substrate are

20   incubated in the presence of labeled polynucleotide or polypeptide. After incubation under conditions for either a nucleic acid or amino acid sequence, the substrate is washed, and any position on the substrate retaining label, which indicates specific binding or complex formation, is assayed, and the binding molecule is identified. Data obtained using different concentrations of the polynucleotide or polypeptide are used to calculate affinity between the labeled nucleic acid or protein and the bound

25   molecule.

What is claimed is:

1. A composition comprising an isolated polynucleotide that is coexpressed with one or more known atherosclerosis-associated genes in a plurality of samples and that is selected from the group consisting of:

5   (a) a nucleic acid sequence selected from SEQ ID NOs:1-34;

           (b) a nucleic acid sequence encoding SEQ ID NO:35;

           (c) a nucleic acid sequence which is the complement of (a) or (b).

2. A polynucleotide comprising the nucleic acid sequence of SEQ ID NO:8 or the complement thereof.

10   3. A composition comprising the polynucleotide of claim 1.

4. A method of using a polynucleotide to screen a library of molecules or compounds to identify at least one ligand which specifically binds the polynucleotide, the method comprising:

           (a) combining the polynucleotide of claim 1 with a library of molecules or compounds under conditions to allow specific binding, and

15   (b) detecting specific binding, thereby identifying a ligand which specifically binds the polynucleotide.

5. The method of claim 4 wherein the library is selected from DNA molecules, RNA molecules, PNAs, mimetics, and proteins.

6. A ligand identified by the method of claim 4 which modulates the activity of the

20 polynucleotide.

7. A method of using a polynucleotide of to purify a ligand which specifically binds the polynucleotide, the method comprising:

           (a) combining the polynucleotide of claim 1 with a sample under conditions to allow specific binding,

25   (b) detecting specific binding between the polynucleotide and a ligand,

           (c) recovering the bound polynucleotide, and

           (d) separating the polynucleotide from the ligand, thereby obtaining purified ligand.

8. A method for diagnosing a disease or condition associated with the altered expression of a polynucleotide that is coexpressed with one or more known atherosclerosis-associated genes in a

30 sample , the method comprising the steps of:

           (a) hybridizing the composition of claim 1 to a sample under conditions to form one or more hybridization complexes;

           (b) detecting the hybridization complexes; and

           (c) comparing the levels of the hybridization complexes with the level of hybridization

35 complexes in a non-diseased sample, wherein the altered level of hybridization complexes compared

with the level of hybridization complexes of a non-diseased sample indicates the presence of the disease or condition.

9. An expression vector comprising the polynucleotide of claim 2.

10. A host cell comprising the expression vector of claim 9.

5   11. A method for producing the polypeptide, the method comprising:

(a) culturing the host cell of claim 10 under conditions for expression of the polypeptide,

(b) recovering the polypeptide from cell culture.

12. A substantially purified polypeptide comprising the product of a gene that is coexpressed with one or more known atherosclerosis-associated genes in a plurality of samples.

10   13. The polypeptide of claim 12, comprising a polypeptide sequence selected from

(a) the polypeptides encoded by SEQ ID NOs:1-34; and

(b) an oligopeptide sequence comprising at least 6 sequential amino acids of the polypeptide sequence of a).

14. The polypeptide comprising the amino acid sequence of SEQ ID NO:35.

15   15. A pharmaceutical composition comprising a polypeptide of claim 12 and a pharmaceutical carrier.

16. A method for using a polypeptide to screen a library of molecules or compounds to identify at least one ligand which specifically binds the polypeptide, the method comprising:

(a) combining the polypeptide of claim 12 with the library of molecules or compounds

20   under conditions to allow specific binding, and

(b) detecting specific binding between the polypeptide and ligand, thereby identifying a ligand which specifically binds the polypeptide.

17. The method of claim 16 wherein the library is selected from DNA molecules, RNA molecules, PNAs, mimetics, proteins, agonists, antagonists, and antibodies.

25   18. A ligand identified by the method of claim 16 which modulates the activity of the polypeptide.

19. A method of using the polypeptide to purify a ligand from a sample, the method comprising:

(a) combining the polypeptide of claim 12 with a sample under conditions to allow specific

30   binding,

(b) detecting specific binding between the polypeptide and a ligand,

(c) recovering the bound polypeptide, and

(d) separating the polypeptide from the ligand, thereby obtaining purified ligand.

20. A method for treating a disease associated with the altered expression of a gene that is

35   coexpressed with one or more known atherosclerosis-associated genes in a subject in need, the method

comprising the step of administering to the subject in need the pharmaceutical composition of claim 15 in an amount effective for treating the disease.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
      JONES, Karen Anne
      VOLKMUTH, Wayne
      WALKER, Michael

<120> ATHEROSCLEROSIS-ASSOCIATED GENES

<130> PB-0013 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/349,015
<151> 1999-07-07

<160> 35

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 1334
<212> DNA
<213> HOMO SAPIENS
<222> 674, 735, 788

<400> 1
 aggcctccct ccacctgtct tctcagagca gataatggca agcatggctg ccgtgctcac   60
 ctgggctctg gctcttcttt cagcgttttc ggccacccag gcacggaaag gcttctggga   120
 ctacttcagc cagaccagcg gggacaaagg cagggtggag cagatccatc agcagaagat   180
 ggctcgcgag cccgcgaccc tgaaagacag ccttgagcaa gacctcaaca atatgaacaa   240
 gttcctggaa aagctgaggc ctctgagtgg gagcgaggct cctcggctcc cacaggaccc   300
 ggtgggcatg cggcggcagc tgcaggagga gttggaggag gtgaaggctc gcctccagcc   360
 ctacatggca gaggcgcacg agctggtggg ctggaatttg gagggcttgc ggcagcaact   420
 gaagccctac acgatggatc tgatggagca ggtggccctg cgcgtgcagg agctgcagga   480
 gcagttgcgc gtggtggggg aagacaccaa ggcccagttg ctggggggcg tggacgaggc   540
 ttgggctttg ctgcagggac tgcagagccg cgtggtgcac cacaccggcc gcttcaaaga   600
 gctcttccac ccatacgccg agagcctggt gagcggcatc gggcgccacg tgcaggagct   660
 gcaccgcagt gtgntccgca cgccccccgcc agccccgcgc gcctcagtcg ctgcgtgcag   720
 gtgctctccc ggaantcacg ctcaaggcca aggccctgca cgcacgcatc cagcagaacc   780
 tggaccantg cgcgaagagc tcagcagagc ctttgcaggc actgggactg aggaaggggc   840
 cggcccggac ccccagatgc tctccgagga ggtgcgccag cgacttcagg ctttccgcca   900
 ggacacctac ctgcagatag ctgccttcac tcgcgccatc gaccaggaga ctgaggaggt   960
 ccagcagcag ctggcgccac ctccaccagg ccacagtgcc ttcgccccag agtttcaaca   1020
 aacagacagt ggcaaggttc tgagcaagct gcaggcccgt ctggatgacc tgtgggaaga   1080
 catcactcac agccttcatg accagggcca cagccgtctg ggggacccct gaggatctac   1140
 ctgcccaggc ccattcccag cttcttgtct ggggagcctt ggctctgagc ctctagcatg   1200
 gttcagtcct tgaaagtggc ctgttgggtg gagggtggaa ggtcctgtgc aggacaggga   1260
 ggccaccaaa ggggctgctg tctcctgcat atccagcctc ctgcgactcc ccaatgcagg   1320
 atgcattcat tcac   1334

<210> 2
<211> 1702
<212> DNA
<213> HOMO SAPIENS

<400> 2
 cgttcccact gcaccctgga gaacgagcct ttgcggggtt tctcctggct gtcctccgac   60
 cccggcggtc tcgaaagcga cacgctgcag tgggtggagg agccccaacg ctcctgcacc   120

```
gcgcggagat gcgcggtact ccaggccacc ggtggggtcg agcccgcagg ctggaaggag      180
atgcgatgcc acctgcgcgc caacggctac ctgtgcaagt accagtttga ggtcttgtgt      240
cctgcgccgc gccccggggc cgcctctaac ttgagctatc gcgcgccctt ccagctgcac      300
agcgccgctc tggacttcag tccacctggg accgaggtga gtgcgctctg ccggggacag      360
ctcccgatct cagttacttg catcgcggac gaaatcggcg ctcgctggga caaactctcg      420
ggcgatgtgt tgtgtccctg ccccggggagg tacctccgtg ctggcaaatg cgcagagctc      480
cctaactgcc tagacgactt gggaggcttt gcctgcgaat gtgctacggg cttcgagctg      540
gggaaggacg gccgctcttg tgtgaccagt ggggaaggac agccgaccct gggggggacc      600
ggggtgccca ccaggcgccc gccggccact gcaaccagcc ccgtgccgca gagaacatgg      660
ccaatcaggg tcgacgagaa gctgggagag acaccacttg tccctgaaca agacaattca      720
gtaacatcta ttcctgagat tcctcgatgg ggatcacaga gcacgatgtc tacccttcaa      780
atgtcccttc aagccgagtc aaaggccact atcaccccat cagggagcgt gatttccaag      840
tttaattcta cgacttcctc tgccactcct caggctttcg actcctcctc tgccgtggtc      900
ttcatatttg tgagcacagc agtagtagtg ttggtgatct tgaccatgac agtactgggg      960
cttgtcaagc tctgctttca cgaaagcccc tcttcccagc caaggaagga gtctatgggc     1020
ccgccgggcc tggagagtga tcctgagccc gctgctttgg gctccagttc tgcacattgc     1080
acaaacaatg gggtgaaagt cggggactgt gatctgcggg acagagcaga gggtgccttg     1140
ctggcggagt cccctcttgg ctctagtgat gcatagggaa acaggggaca tgggcactcc     1200
tgtgaacagt ttttcacttt tgatgaaacg gggaaccaag aggaacttac ttgtgtaact     1260
gacaatttct gcagaaatcc cccttcctct aaattccctt tactccactg aggagctaaa     1320
tcagaactgc acactccttc cctgatgata gaggaagtgg aagtgccttt aggatggtga     1380
tactggggga ccgggtagtg ctggggagag atattttctt atgtttattc ggagaatttg     1440
gagaagtgat tgaacttttc aagacattgg aaacaaatag aacacaatat aatttacatt     1500
aaaaaataat ttctaccaaa atggaaagga aatgttctat gttgttcagg ctaggagtat     1560
attggttcga aatcccaggg aaaaaaataa aaataaaaaa ttaaaggatt gttgataaaa     1620
aaaaaaaaaa aaaagatct ttaattaagc ggcccaagct tattcccttt agtgagggtt     1680
aattttagct tgcactggcc ac                                             1702
```

<210> 3
<211> 586
<212> DNA
<213> HOMO SAPIENS
<222> 48, 66, 560, 574, 577, 580

<400> 3
```
tcgaggactc cgccaactac agctgcgtct acgtggacct gaagccgnct ttcggggggct      60
acgcgnccag cgagcgcttg gagctgcacg tggacggacc ccctcccagg cctcagctcc      120
gggcgacgtg gagtggggcg gtcctggcgg gccgagatgc cgtcctgcgc tgcgagggac      180
ccatccccga cgtcaccttc gagctgctgc gcgagggcga dacgaaggcc gtgaagacgg      240
tccgcacccc cggggccgcg gcgaacctcg agctgatctt cgtgggggccc cagcacgccg      300
gcaactacag gtgccgctac cgctcctggg tgccccacac cttcgaatcg gagctcagcg      360
accctgtgga gctcctggtg gcagaaagct gatgcagccg cgggcccagg gtgctgttgg      420
tgtcctcaga agtgccgggg attctggact ggctccctcc cctcctgttg cagcacaagg      480
ccggggtctc tggggggctg gagaagcctc cctcattcct cccaggaatt aataaatgtg      540
aagagagctc tgtttaaaan aaaaaaaaag aaanaanaan aaccaa                    586
```

<210> 4
<211> 433
<212> DNA
<213> HOMO SAPIENS

<400> 4
```
ctcaagaccc agcagtggga cagccagaca gacggcacga tggcactgag ctcccagatc      60
tgggccgctt gcctcctgct cctcctcctc ctcgccagcc tgaccagtgg ctctgttttc      120
ccacaacaga cgggacaact tgcagagctg caaccccagg acagagctgg agccagggcc      180
agctggatgc ccatgttcca gaggcgaagg aggcgagaca cccacttccc catctgcatt      240
ttctgctgcg gctgctgtca tcgatcaaag tgtgggatgt gctgcaagac gtagaaccta      300
cctgccctgc ccccgtcccc tcccttcctt atttattcct gctgccccag aacataggtc      360
ttggaataaa atggctggtt cttttgtttt ccaaaaaaaa aaaaaaaaaa aaaaaaaaaa      420
aaaaaaaaaa aaa                                                        433
```

<210> 5
<211> 752

```
<212> DNA
<213> HOMO SAPIENS

<400> 5
  attgtacact ttaaaataat ggaattttac agtaagtgaa gtatgtatcg atgaagctat          60
  taaaacatct atttattagc tcaaatttct acaggtcaga attctggcat ggagtggctg         120
  cattctttgt ttaagctgaa atcaaggtgt tgtttggccg tgttctcacc tgaagctcag         180
  agttcacttc caagctcatt tttgtcctta gcagaattga gtgtcttgca attgtagaac         240
  tgaggtcttg gcttgctgtc tgtcagcagg ggactgctcc ctgcttctag aggccaccgg         300
  tttccctcgt tatgtggccc cttccatttt caggccagca ataatgtgtt gaatacttcc         360
  tatgcttcaa atctctggct tctgctacca gctggagaaa aaactctctg cttgtagagg         420
  gctcatgtga tttacttagt ctttgtctta aggtcaattt atttggtact tgggatttta         480
  attgtatctg tatgtttcca tcaaggcaat aactgtatta gtgtttgaat aaataaccag         540
  gtaatctggt aatttaccat actggtaatc tgacagggag atgggaattc atctttataa         600
  ttctgcttac cacaaaccat gtctgtgctt attttctttg gggaagagtt gtctgtgact         660
  gtcacttagt ttgaggttcc atgttgctga gattctgtcc agtattttga cctcttcccc         720
  aaatctggtc ttcagaacca tctcttagga gc                                       752

<210> 6
<211> 944
<212> DNA
<213> HOMO SAPIENS

<400> 6
  tcttgggccg agaatttttt tttttttttt tttgcttggt cgggtaattt tcattccaaa          60
  taaacttatc acaaaaaaac tcagcttccc aaggtcattc ccccgctgcc agatacatac         120
  ttatctctga aagagtttgg aagatggacc tttcaattcc tctacaatta gtagctgagt         180
  tacagagtaa cctgccagca atcctatcag cattcatcag actatttaaa tagagcaaag         240
  tccacaaaaa gttccactga gacatgctga gcaaaggccg gagccccaga agaaaacaag         300
  tacagactca gaggaaagct gccctggtcc tgagtgtgac tcccatggtc cccgtggggt         360
  ctgtgtggtt ggcaatgagc tctgtgctgt cagctttcat gagggagctc cctggctggt         420
  tcctgttctt tggggtcttc ctccccatga ctttgctgct gctcctcctc atcgcctact         480
  tcaggatcaa actgattgag gttaatgaag aactgtccca gaactgtgat cgccaacata         540
  atcccaagga tggctcttcc ctgtaccaga gaatgaaatg gacgtgaagt tggtgacttt         600
  ccaataacta aagcacaatg agtttctact ggtcagcaag caatggccaa cagttcagct         660
  aataaagtag gttgataaac tagaaccata gcaaaataga aagaatacta agatactcat         720
  tctgaaccat actgaaaagt ggcagctatt atctaagggg acttctcaga gactcagtat         780
  aacagcagct cttgaaaagt accaagaatg gatttcctgg gtatatacac tggacacatt         840
  gtaacttttt aacttttatt gtgactgtgt ctgctctaaa cggcatattt aaaaaataaa         900
  attctgcagc atcttactac ataaaaaaaa aaaaaaaaa aaaa                           944

<210> 7
<211> 868
<212> DNA
<213> HOMO SAPIENS

<400> 7
  cctccctccg cgagctggac gctccgcagc ccgcccgcca gccggcccgc cggccgccgc          60
  aggaatccct ggataaagac cagctcaacc atcgctgaga aaacagacct aggcttccca         120
  gggcggttaa cccgccggcc tctgggcaga gactaaaaga caaaacaaaa taaaacaaca         180
  acaaaaaact cccagtgtgt ttcctactct tctttgtctt ggaggaaagc aaagggagag         240
  aaatggactt caccagtggt ctttggcttc atcaattcac aggaaatggc atcaagatgg         300
  ttcaactaag acatgatcac taaaaacatt ataataatac ctttttgaaa aactcagttt         360
  ctcctgttta ctaaatattt atttcatcaa catgggctgc gttccactgt gtcaggattc         420
  tgcatgtggg tggagcactg ttccagcctg agaagatggt tctgaggcca cttagcaaga         480
  cattttccag catgagcagg tttctctgtg gaaatagtga cacctgttct ggtgtgttgt         540
  ctttcctcag ggaacttaag gggtacaaag ctcctgaaaa tgttctttat gctggttgaa         600
  gctcttatgt cgctgtactg attccctacg atgcagattt gaatcacaga gtaattaaaa         660
  tatggatcaa ataaggctgg ggctcaccaa ggctgaaagc tgtagccatt caaggcatca         720
  tttctgtcat gaaaatatag gaccttttca aacatgcct tcaggaaggt gttctctttt          780
  caaacaaaag tctaatgact gcataactct tcttgaccac atcttacact ttctctagac         840
  ttgcttattt acagctactg gaacaaaa                                            868
```

```
<210> 8
<211> 3111
<212> DNA
<213> HOMO SAPIENS
<222> 44

<400> 8
  cgagggcgga cgcaaagaac gcggaggacc tctgggtgcc tgcnggggag ctgctccagc     60
  cgggccgccg ggagcggtgg ggagagcatc gcgcagccgc ccctccacgc gcccgcccag    120
  ccgcgttcgc ccactgggct ctcccggctg cagtgccagg gcgcaggacg cggccgatct    180
  cccgctcccg ccacctccgc caccatgctg ctcccccagc tctgctggct gccgctgctc    240
  gctgggctgc tcccgccggt gcccgctcag aagttctcgg cgctcacgtt tttgagagtg    300
  gatcaagata aagacaagga ttgtagcttg gactgtgcgg gttcgcccca gaaacctctc    360
  tgcgcatctg acggaaggac cttcctttcc cgttgtgaat ttcaacgtgc caagtgcaaa    420
  gatccccagc tagagattgc atatcgagga aactgcaaag acgtgtccag gtgtgtggcc    480
  gaaaggaagt atacccagga gcaagcccgg aaggagtttc agcaagtgtt cattcctgag    540
  tgcaatgacg acggcaccta cagtcaggtc cagtgtcaca gctacacggg atactgctgg    600
  tgcgtcacgc ccaacgggag gcccatcagc ggcactgccg tggcccacaa gacgccccgg    660
  tgcccgggtt ccgtaaatga aaagttaccc caacgcgaag gcacaggaaa aacagatgat    720
  gccgcagctc cagcgttgga gactcagcct caaggagatg aagaagatat tgcatcacgt    780
  tacccctaccc tttggactga acaggttaaa agtcggcaga acaaaaccaa taagaattca    840
  gtgtcatcct gtgaccaaga gcaccagtct gccctggagg aagccaagca gcccaagaac    900
  gacaatgtgg tgatccctga gtgtgcgcac ggcggcctct acaagccagt gcagtgccac    960
  ccctccacgg ggtactgctg gtgcgtcctg gtggacacgg ggcgccccat tcccggcaca   1020
  tccacaaggt acgagcagcc gaaatgtgac aacacgggcc agggcccacc cagccaaagc   1080
  ccgggacctg tacaagggcc gccagctaca aggttgtccg ggtgccaaaa agcatgagtt   1140
  tctgaccagc gttctggacg cgctgtccac ggacatggtc cacgccgcct ccgacccctc   1200
  ctcctcgtca ggcaggctct cagaacccga ccccagccat accctagagg agcgggtggt   1260
  gcactggtac ttcaaactac tggataaaaa ctccagtgga gacatcggca aaaggaaat    1320
  caaacccttc aagaggttcc ttcgcaaaaa atcaaagccc aaaaaatgtg tgaagaagtt   1380
  tgttgaatac tgtgacgtga ataatgacaa atccatctcc gtacaagaac tgatgggctg   1440
  cctgggcgtg gcgaaagagg acggcaaagc ggacaccaag aaacgccaca ccccagagg    1500
  tcatgctgaa agtacgtcta atagacagcc aaggaaacaa ggataaatgg ctcataccccc   1560
  gaaggcagtt cctagacaca tgggaaattt ccctcaccaa agagcaatta agaaaacaaa   1620
  aacagaaaca catagtattt gcactttgta ctttaaatgt aaattcactt tgtagaaatg   1680
  agctatttaa acagactgtt ttaatctgtg aaaatggaga gctggcttca gaaaattaat   1740
  cacataccaa tgtatgtgtc ctctttgac cttggaaatc tgtatgtggt ggagaagtat   1800
  ttgaatgcat ttaggcttaa tttcttcgcc ttccacatgt taacagtaga gctctatgca   1860
  ctccggctgc aatcgtatgg ctttctctaa cccctgcagt cacttccaga tgcctgtgct   1920
  tacagcattg tggaatcatg ttggaagctc cacatgtcca tggaagtttg tgatgtacgg   1980
  ccgaccctac aggcagttaa catgcatggg ctggtttgtt tcttgggatt ttctgttagt   2040
  ttgtcttgtt ttgctttcca gagatcttgc tcatacaatg aatcacgcaa ccactaaagc   2100
  tatccagtta agtgcaggta gttcccctgg aggaaataat attttcaaac tgtcgttggt   2160
  gtgatacttt ggctcaaagg atctttgctt ttccatttta agcttctgtt ttgagttttg   2220
  ccctggggct tgaatgagtc ccagagagtc gttcggatgg tgggaggctg cctaggaggc   2280
  agtaaatcca gttcacagtg cctgggaggg gcccatcctt ccaaaatgta aatccagttc   2340
  gcggtgtgac cgagctgggc taacaggctt gtctgccggc ttttcctacc tacacgtgga   2400
  cattattctc ctgatcctcc tacctggttc caccccaggg ctaccggaag gtaaaatctt   2460
  cacctgaacc aattatgagc agtctcctta ctgaaggtac agccggatac gtggtgcccc   2520
  cggggctggt gttggcagcc gggggggaggt gcctgagggt ccccacggtt cctttctgct   2580
  tttctgaatg catcaagggt acgagaactt gccaatggga aattcatccg agtggcactg   2640
  gcagagaagg ataggagtgg aatgcccaca cagtgaccaa cagaactggt ctgcgtgcat   2700
  aaccagctgc caccctcagg cctgggcccc agagctcagg gcacccagtg tcttaaggaa   2760
  ccatttggag gacagtctga gagcaggaac ttcaagctgt gattctatct cggctcagac   2820
  ttttggttgg aaaaagatct tcatggcccc aaatcccctg agacatgcct tgtagaatga   2880
  ttttgtgatg ttgtgatgct tgtggagcat cgcgtaaggc ttcttgctta tttaaactgt   2940
  gcaaggtaaa aatcaagcct tggagccac agaaccagct caagtacatg ccaatgttgt   3000
  ttaagaaaca gttatgatcc taaacttttt ggataatctt ttatatttct gacctttgaa   3060
  tttaatcatt gttcttagat taaaataaaa tatgctattg aaactaaaaa a            3111
```

```
<210> 9
<211> 2311
<212> DNA
```

<213> HOMO SAPIENS
<222> 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487,
488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502,
503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517,
518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532,
533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 2288, 2295,
2296, 2297, 2298, 2299

<400> 9
```
gccgctcgcc cacggactcc gacgtgtccc tcgactccga ggactccggg gctaagtctc    60
caggcatcct gggctacaat atctgtcccc gcgggtggaa tggcagcctt cggctcaagc   120
gtggcagcct ccccgccgag gcctcctgca ccacctagag ccccacccc gaccccaccc   180
cgggagggca gagccagaag aaggctcatt agacctgggg gacccaaagg gtctggcctc   240
tttgggcagc cccagagatg aggggtcagc agaggagagc tctgggggttg gggatgggtt   300
agggacgcaa gcttgagttc tagcccttgc tctcattcag ctgttgtgtg accctgggta   360
agacccttcc ttgtttgacc ctcagctttc ccatctgttt aatggtggct ttggccaagg   420
caatccacaa acgtcaaaat tcccttccc atcagtacac acaccgatgc acannnnnnn   480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   540
nnnntagtta gtgccttgga tgaggcgggg cagtgtgtat atggacccct ggacttgcta   600
ccttcagggt tccatactcg tccctcccct cctggctctg ctgtctggag tctggcaagc   660
ggggtgtgtt cagaaggtcc taggcctgtg tcgcatgtcc aggcactggc ctgaccatcc   720
ggctccctgg gcaccaagtc ccagggcagg agcagctgtt ttccatccct tcccagacaa   780
gctctatttt tatcacaatg acctttagag aggtctccca ggccagctca aggtgtccca   840
ctatcccctc tggagggaag aggcaggaaa attctccccg ggtccctgtc atgctacttt   900
ctccatccca gttcagactg tccaggacat cttatctgca gccataagag aattataagg   960
cagtgatttc ccttaggccc aggacttggg cctccagctc atctgttcct tctgggccca  1020
ttcatgggca ggttctgggc tcaaagctga actggggaga gaagagatac agagctacca  1080
tgtgactttа cctgattgcc ctcagtttgg ggttgcttat gggaaagag agagacaaag  1140
agttacttgt tacgggaaat atgaaaagca tggccaggat gcatagagga gattctagca  1200
ggggacagga ttggctcaga tgacccctga gggctcttcc agtcttgaaa tgcattccat  1260
gatattagga agtcgggggt gggtggtggt ggtgggctag ttgggcttga atttaggggc  1320
cgatgagctt gggtacgtga gcagggtgtt aagttagggt ctgcctgtat ttctggtccc  1380
cttgggaaat gtcccttct tcagtgtcag acctcagtcc cagtgtccat atcgtcccca  1440
gaaaagtaga cattatcctg ccccatccct tccccagtgc actctgacct agctagtgcc  1500
tggtgcccag tgacctgggg gagcctggct gcaggccctc actggttccc taaaccttgg  1560
tggctgtgat tcaggtcccc agggggggact cagggaggaa tatggctgag ttctgtagtt  1620
tccagagttg ggctggtaga gctttctaga ggttcagaat attagcttca ggatcagctg  1680
ggggtatgga attggctgag gatcaaacgt atgtaggtga aaggatacca ggatgttgct  1740
aaaggtgagg gacagtttgg gtttgggact taccggggtg atgttagatc tggaacccc  1800
aagtgaggct ggagggagtt aaggtcagta tggaagatag ggttgggaca gggtgctttg  1860
gaatgaaaga gtgacctag aggggctcctt gggcctcagg aatgctcctg ctgctgtgaa  1920
gatgagaagg tgctcttact cagttaatga tgagtgacta tatttaccaa agcccctacc  1980
tgctgctggg tcccttgtag cacaggagac tggggctaag ggcccctccc agggaaggga  2040
caccatcagg cctctggctg aggcagtagc atagaggatc catttctacc tgcatttccc  2100
agaggactag caggaggcag ccttgagaaa ccggcagttc ccaagccagc gcctggctgt  2160
tctctcattg tcactgccct ctccccaacc tctcctctaa cccactagag attgcctgtg  2220
tcctgcctct tgcctcttgt agaatgcagc tctggccctc aataaatgct cctgcattc   2280
taaaaaanaa aaaannnnna aaaaaaaag g                                  2311
```

<210> 10
<211> 1866
<212> DNA
<213> HOMO SAPIENS

<400> 10
```
agcttttgtt cacactttaa atagcagtcc cagaatgatt tcactacaga ctctctggaa    60
agcctgggag ctgaattccg gaagatcccc acatcgatga aagcaaagcg aagccaccaa   120
gccatcatca tgtccacgtc gctacgagtc agcccatcca tccatggcta ccacttcgac   180
acagcctctc gtaagaaagc cgtgggcaac atctttgaaa acacagacca agaatcacta   240
gaaaggctct tcagaaactc tggagacaag aaagcagagg agagagccaa gatcatttt   300
gccatagatc aagatgtgga ggagaaaacg cgtgccctga tggccttgaa gaagaggaca   360
aaagacaagc ttttccagtt tctgaaactg cggaaatatt ccatcaaagt tcactgaaga   420
gaagaggatg gataaggacg ttatccaaga atggacattc aaagaccaag tgagtttgtg   480
```

```
agattctaac agatgcagca ttttgctgct accttacaag cttctcttct gtcaggactc      540
cagaggctgg aaagggaccg ggactggaaa gggaccagga ctgaacagac tggttacaaa      600
gactccaaac aatttcatgc cctgtgctgt tacagaggag aacaaaatgc tttcagcaag      660
gatttgaaaa ctcttccgtc cctgcaggaa aggattgatg ctgatagaag agcctggaca      720
gatgtaatga gaactaaaga aaacagatgg ctggagatga catttatcca gggtcacttt      780
gtcaggccct aggacttaaa tcgaagttga acttttttt tttttaacc aaatagatag      840
gggaagggag gagggagagg gaggacaggg agagaaaata ccatgcataa attgtttact      900
gaattttat atctgagtgt tcaaaatatt tccaagcctg agtattgtct attggtatag      960
attttagaa atcaataatt gattatttat ttgcacttat tacaatgcct gaaaaagtgc     1020
accacatgga tgttaagtag aaattcaaga aagtaagatg tcttcagcaa ctcagtaaaa     1080
ccttacgcca ccttttggtt tgtaaaaggt tttttataca tttcaaacag gttgcacaaa     1140
agttaaaata atggggtctt ttataaatcc aaagtactgt gaaaacattt tacatatttt     1200
ttaaatcttc tgactaatgc taaaacgtaa tctaattaaa tttcatacag ttactgcagt     1260
aagcattagg aagtgaatat gatatacaaa atagtttata aagactctat agtttctata     1320
atttatttta ctggcaaatg tcatgcaaca ataataaatt attgtaaact ttgtggcttt     1380
tggtctgtga tgcttggtct caaaggaaaa aataagatgg taaatgttga tatttacaaa     1440
cttttctaaa gatgtgtctc taacaataaa agttaatttt agagtagttt tatattaatt     1500
accaaacttt ttcaaaacaa attcttacgt caaatatctg ggaagtttct ctgtcccaat     1560
cttaaaatat aaaatataga tatagaagtt catagattga ctccttggca tttctattta     1620
tgtatccatt aaggatgagt tttaaaaggc tttctcttca tacttttgaa aaatttcttc     1680
tatgattaca gtagctatgt acatgtgtac atctattttt cccaagcaat atgtttnggg     1740
tttagagtct gagtgatgac caagattctg tgtgttacta ctgtttgttt aataggaaca     1800
aatatagaaa taatattatc tctttgctta tttcccgtta aaactataat aaaatgtttc     1860
taggaa                                                              1866
```

<210> 11
<211> 1929
<212> DNA
<213> HOMO SAPIENS

<400> 11

```
gctgcctgcc ggtgctcttc gtggctctgg gcatggcctc ggaccccatc ttcacgctgg      60
cgccccgct gcattgccac tacgggcct tcccccctaa tgcctctggc tgggagcagc      120
ctcccaatgc cagcggcgtc agcgtcgcca gcgctgccct agcagccagc gccgccagcc      180
gtgtcgccac cagtaccgac ccctcgtgca gcggcttcgc cccgccggac ttcaaccatt      240
gccctcaagg attgggacta taatggcctt cctgtgctca ccaccaacgc catcggccag      300
tgggatctgg tgtgtgacct gggctggcag gtgatcctgg agcagatcct cttcatcttg      360
ggctttgcct ccggctacct gttcctgggt taccccgcag acagatttgg ccgtcgcggg      420
attgtgctgc tgaccttggg gctggtgggc ccctgtggag taggaggggc tgctgcaggc      480
tcctccacag gcgtcatggc cctccgattc ctcttgggct ttctgcttgc cggtgttgac      540
ctgggtgtct acctgatgcg cctggagctg tgcgacccaa cccagaggct tcgggtggcc      600
ctggcagggg agttggtggg ggtgggaggg cacttcctgt cctgggcct ggcccttgtc      660
tctaaggatt ggcgattcct acagcgaatg atcaccgctc cctgcatcct cttcctgttt      720
tatggctggc ctggtttgtt cctggagtcc gcacggtggc tgatagtgaa gcggcagatt      780
gaggaggctc agtctgtgct gaggatcctg gctgagcgaa accggcccca tgggcagatg      840
ctgggggagg aggcccagga ggccctgcag gacctggaga atacctgccc tctccctgca      900
acatcctcct tttcctttgc ttccctcctc aactaccgca acatctggaa aaatctgctt      960
atcctgggct tcaccaactt cattgcccat gccattcgcc actgctacca gcctgtggga     1020
ggaggaggga gcccatcgga cttctacctg tgctctctgc tggccagcgg caccgcagcc     1080
ctggcctgtg tcttcctggg ggtcaccgtg gaccgatttg ccgccggggg catccttctt     1140
ctctccatga cccttaccgg cattgcttcc ctggtcctgct tgggctct tctcctccca agctgcgcc     1200
aacgaggctg ccatccacac tttctctgtc cttgggctct tctcctccca agctgcgcc     1260
atcctcagca ccctccttgc tgctgaggtc atccccacca ctgtccgggg ccgtggcctg     1320
ggcctgatca tggctctagg ggcgcttgga ggactgagcg gcccggccca gcgcctccac     1380
atgggccatg gagccttcct gcagcacgtg gtgctggcgg cctgcgccct cctctgcatt     1440
ctcagcatta tgctgctgcc ggagaccaag cgcaagctcc tgcccgaggt gctccgggac     1500
ggggagctgt gtcgccggcc ttccctgctg cggcagccac cccctacccg ctgtgaccac     1560
gtcccgctgc ttgccacccc caaccctgcc ctctgagcgg cctctgagta ccctggcggg     1620
aggctggccc acacagaaag gtggcaagaa gatcgggaag actgagtagg gaaggcaggg     1680
ctgcccagaa gtctcagagg cacctcacgc cagccatcgc ggagagctca gaggccgtc     1740
cccaccctgc ctcctccctg ctgctttgca ttcacttcct tggccagagt caggggacag     1800
ggagagagct ccacactgta accactgggt ctgggctcca tcctgcgccc aaagacatcc     1860
acccagacct cattatttct tgctctatca ttctgtttca ataaagacat ttggaataaa     1920
```

                                                                      1929
     aaaaaaaaa

<210> 12
<211> 1831
<212> DNA
<213> HOMO SAPIENS

<400> 12
   ctggagccgc cctgggtgtc agcggctcgg ctcccgcgca cgctccggcc gtcgcgcacc   60
   tcgggcacct gcaggtccgt ggcgtcccgc ggctgggcgc ccctgactcc gtcccggcca   120
   gggagggcca tgatttccct cccgggggccc ctggtgacca acttgctgcg gttttttgttc   180
   ctggggctga gtgccctcgc gcccccctcg cgggcccagc tgcaactgca cttgcccgcc   240
   aaccggttgc aggcggtgga gggagggggaa gtggtgcttc cagcgtggta caccttgcac   300
   ggggaggtgt cttcatccca gccatgggag gtgccctttg tgatgtggtt cttcaaacag   360
   aaagaaaagg aggatcaggt gttgtcctac atcaatgggg tcacaacaag caaacctgga   420
   gtatccttgg tctactccat gccctcccgg aacctgtccc tgcggctgga gggtctccag   480
   gagaaagact ctggcccccta cagctgctcc gtgaatgtgc aagacaaaca aggcaaatct   540
   aggggccaca gcatcaaaac cttagaactc aatgtactgg ttcctccagc tcctccatcc   600
   tgccgtctcc agggtgtgcc ccatgtgggg gcaaacgtga ccctgagctg ccagtctcca   660
   aggagtaagc ccgctgtcca ataccagtgg gatcggcagc ttccatcctt ccagactttc   720
   tttgcaccag cattagatgt catccgtggg tctttaagcc tcaccaacct ttcgtcttcc   780
   atggctggag tctatgtctg caaggcccac aatgaggtgg gcactgccca atgtaatgtg   840
   acgctggaag tgagcacagg tcagtgaggg ggcctggagc tgcagtggtt gctggagctg   900
   ttgtgggtac cctggttgga ctggggttgc tggctgggct ggtcctcttg taccaccgcc   960
   ggggcaaggc cctggaggag ccagccaatg atatcaagga ggatgccatt gctccccgga   1020
   ccctgccctg gcccaagagc tcagacacaa tctccaagaa tgggaccctt tcctctgtca   1080
   cctccgcacg agccctccgg ccaccccatg gccctcccag gcctggtgca ttgacccccca   1140
   cgcccagtct ctccagccag gccctgccct caccaagact gcccacgaca gatggggccc   1200
   accctcaacc aatatccccc atccctggtg gggtttcttc ctctggcttg agccgcatgg   1260
   gtgctgtgcc tgtgatggtg cctgcccaga gtcaagctgg ctctctggta tgatgacccc   1320
   accactcatt ggctaaagga tttgggggtct ctccttccta taagggtcac ctctagcaca   1380
   gaggcctgag tcatgggaaa gagtcacact cctgacccctt agtactctgc ccccacctct   1440
   ctttactgtg ggaaaaccat ctcagtaaga cctaagtgtc caggagacag aaggagaaga   1500
   ggaagtggat ctggaattgg gaggagcctc cacccacccc tgactcctcc ttatgaagcc   1560
   agctgctgaa attagctact caccaagagt gaggggcaga gacttccagt cactgagtct   1620
   cccaggcccc cttgatctgt accccacccc tatctaacac caccccttggc tcccactcca   1680
   gctccctgta ttgatataac ctgtcaggct ggcttggtta ggtttttactg gggcagagga   1740
   tagggaatct cttattaaaa ctaacatgaa atatgtgttg ttttcatttg caaatttaaa   1800
   taaagataca taatgtttgt atgagataag a   1831

<210> 13
<211> 909
<212> DNA
<213> HOMO SAPIENS

<400> 13
   gaggaggtgg gcgccaacag acaggcgatt aatgcggctc ttacccaggc aaccaggact   60
   acagtataca ttgtggacat tcaggacata gattctgcag ctcgggcccg acctcactcc   120
   tacctcgatg cctactttgt cttccccaat gggtcagccc tgacccttga tgagctgagt   180
   gtgatgatcc ggaatgatca ggactcgctg acgcagctgc tgcagctggg gctggtggtg   240
   ctgggctccc aggagagcca ggagtcagac ctgtcgaaac agctcatcag tgtcatcata   300
   ggattgggag tggctttgct gctggtcctt gtgatcatga ccatggcctt cgtgtgtgtg   360
   cggaagagct acaaccggaa gcttcaagct atgaaggctg ccaaggaggc caggaagaca   420
   gcagcagggg tgatgccctc agcccctgac atcccaggga ctaacatgta caacactgag   480
   cgagccaacc ccatgctgaa cctccccaac aaagacctgg gcttggagta ccctctctccc   540
   tccaatgacc tggactctgt cagcgtcaac tccctggacg acaactctgt ggatgtggac   600
   aagaacagtc aggaaatcaa ggagcacagg ccaccacaca caccaccaga gccagatcca   660
   gagcccctga gcgtggtcct gttaggacgg caggcaggcg caagtggaca gctggagggg   720
   ccatcctaca ccaacgctgg cctggacacc acggacctgt gacaggggcc cccactcttc   780
   tggaccccctt gaagaggccc taccacaccc taactgcacc tgtctccctg gagatgaaaa   840
   tatatgacgc tgccctgcct cctgctttttg gccaatcacg gcagacaggg gttggggaaa   900
   tattttattt   909

```
<210> 14
<211> 1453
<212> DNA
<213> HOMO SAPIENS
<222> 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916,
917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931,
932, 933, 934, 935

<400> 14
  ggagccaagt gggccctcg  gcctcttcct tcgttccagg cccatgattt tccctacact        60
  tctccctggc ccaggctcca gccacaggca cctctcctgc ccccgcccac cctcctgacc       120
  gcagctccca ggccctggag acctccaggc tttcctgccc tgggcagccc cacctcacag       180
  ccagagtcaa tgccttcatg ggaagggctc ccagccacac ccagagtggc ccaaagctgt       240
  tgaagtcagc atcctttgtc ccatcaggac cctcctgcct cctctccagg cccttgttcg       300
  cctccccacc ctcctcagag gcccggggaa gggaagagca ggtcagtaca gaggttctgt       360
  ctacagggag gggccctggg tctatgcaca gctggagctc tgagccttcc acagcccgtg       420
  tgactgctag agggcagggg tgcagggctc aggggggccg ggctggtcct ttggggctgg       480
  tgttcctacg tcagtcccca cctggggaat aaactccagc ctctcctgct catacagaag       540
  gaactggttg ggtttgcttt atgggatctt tgagaccaaa acagatgctc ctgtttgctg       600
  ggggagggtg tgagcacgga gtatttctgt ccctcgtgaa gtcacgtcac acaggggaga       660
  ggcgaggtcg atggaactgg ccacgcacag gctctggctc tggaaggagg gatgatgagt       720
  gggcgttttc ccggcaggcc cccggggtcc tcagcctcag caacccaggg agaggacaga       780
  aatgaaccga tggttgaggg attgtcacgg gaggaacatg acacccgaag ggactctagg       840
  tgccctcgga gtgccacaca tgcccagacc ttctcacacc cacacaaata ggctctgccg       900
  tgnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnttgtt cacactcaga acccaggaca       960
  gccacagcca ccgcttaggc gaagccactg cagatgcccc tggaatgggc acagcacagc      1020
  cagggcgctc ttccaggcag gcgaggataa cttgagagtt tcctagggca ccagggacag      1080
  agctcagagg cccccgaggt gtgtgtagga ggcggaggcc cgcagagcac agagcaggag      1140
  aagggcttgg gccctggagg agaaagccat tctggacacc aggggacctg gacggagggt      1200
  ccccacagcc cgtgccccac gccgcctgga ggccagaggg gtcagtggcc ctgctgtccc      1260
  ggctccatct tggttctagc cgccacctgt atgaacacag tggcccggct taacgcacta      1320
  acccagcctc tccctgtgtc ccacagggag tagcaagacc caccccacac tgccttcacc      1380
  atctacacca gtgacgccgc tgtgtgtctt agcatggaaa taaataaacc tgaatgcaaa      1440
  aaaaaaaaaa agg                                                        1453
```

```
<210> 15
<211> 443
<212> DNA
<213> HOMO SAPIENS

<400> 15
  gatatagaca acttccagag tcaccagtgt gcaaatggag ccacctgcat tagtcatact        60
  aatggctatt cttgcctctg ttttggaaat tttacaggaa aattttgcag acagagcaga       120
  ttaccctcaa cagtctgtgg gaatgagaag acaaatctca cttgctacaa tggaggcaac       180
  tgcacagagt tccagactga attaaaatgt atgtgccggc caggttttac tggagaatgg       240
  tgtgaaaagg acattgatga gtgtgcctct gatccgtgtg tcaatggagg tctgtgccag       300
  gacttactca acaaattcca gtgcctctgt gatgttgcct ttgctggcga gcgctgcgag       360
  gtggacttgg cagatgactt gatctccgac attttcacca ctattggctc agtgactgtc       420
  gccttgttac tgatcctctt gct                                               443
```

```
<210> 16
<211> 1537
<212> DNA
<213> HOMO SAPIENS
<222> 284, 285, 287

<400> 16
  aaaaaaaaca acccggtagc attgtccctt ccccactgac aaacttatca aatccagaag        60
  ctttagagtt tcgtctctaa ttatttttct cctgaacaaa attacccaag tcaaaacaaa       120
  atgtattttt agaattacgg cagcatacga cctgaatttt gtgagtttcg tggctttatc       180
  ttaaatcacc atttccctaa aaacggtttc tttctcctta gaatgctgg tggcaacttg        240
  atgaaacagc caaatgcacc agggcaggtc actttcccaa aaannanaag aaaaaaaact       300
  cattgagata gctacagttc tataggttaa tttaaagcct cctttttcta ctcatttttg       360
```

```
aaagcaaaat tacattttac tattttacat aaccagtgaa aagacgttga aagcctacag    420
ctcactgttt ttggtgctct ggaaatgttg agggtgggtt tttaaccagt gattttttaac   480
gtgcagtgaa tttgttagac ttttaaacac cagctaaggt agtcaaactt gatccccatt    540
aaaaatcaag gaattagggg tcggggggagg gtttaggagt gatccagaat gacctcccag    600
aattactgtg cgtacaactt tattttttcag agttttcatt ggaatggtaa gagtttttatg   660
aaagacagtt ttaaaactta ttctgagtta aatattaata ctttaaaaaa ttattgtact    720
agacttattg cagcctttttg aaagtagcag agtttcatca taccacatat ataacagagc   780
ataaatttttc tataatcagg cacctttttgc tgctttttgag taagactgtt ttcctgttta   840
ggtgttaagc atcgccagac ataaaaatct attctctcct ctcgattgta gcatagcctg    900
acagctctag atacagcatt tctatgatga aaaatgagta tccatcagga aatctagaag    960
actagccgtg tttctcaga ctccacctttt gtttgcactc tgttgcctgt gaggagcttt    1020
ctggcatgtg attatttact tcaaaactag agttccaagc acctacatta attattttat    1080
attgtgtgca gaatagtata tctttttaatg tcagatatga tacactgcac atattgcttt   1140
tgcactctta aaatttttgt actaaataat agaaaatatt tatattcttt gagtgtgagc    1200
tttgaataga tggcattatc acttttattgt tttttttaaca aaaacttttt ctcaattatt    1260
ctattgcaat gttattctga gcaagtccta tgccaaatat cttgtataat gtttgtatgg    1320
aagattaaat tttactcttg tgtggtaaga ctatttcagt tactgattttt atagttggaa   1380
tttgatattc cagcacaaag tccacagtgt attcagaaat ccaagttggt gtcatacatt    1440
tcattttgat gtgaactttt ctttgctttc ctttgttcta agactccatt ttgcaataaa    1500
cgttttgaca gtaaaaaaaa taaaaaagga aaaaaaa                             1537
```

<210> 17
<211> 972
<212> DNA
<213> HOMO SAPIENS

<400> 17
```
acgcaaattc ggcacgaggg ttctaaaacc cagtttggtt tacgttgtct ttcacagtag     60
tatatttagc tcttctctgg aaagttgtgg gttaatataa ttcttaaaca tgaaaatgta    120
attaaacaca ccacgagaga acaatattcc aggagactta atagtgatta ctttcttcaa    180
tcaggaaatc gtttcagtgc ctcctttgta ggaatgcttt gtttttgtgat gggtttttctt    240
aaagaagagc acacctccgt ccaatctcct gagacagcca cgtctccgct gacatcccac    300
tgtgatgctt tcagatagtc agtgaatgtt tctgataacc ttcatccagt atctgaaaca    360
caatgtgaga gattatattg ttttagataa taacatccca tttagttgac taaaatcttc    420
caaactctga aagctgcaca ctgctactcc agagagtgca ggtcttagct cttctcctt     480
ctgacttcaa gatgaatctt tgggacgatg tttctggtgc ttggtccaca gtgattcact    540
tttgaaggag aggccacatg acatgaactg cctggtgtta caacctagct aacatatttg    600
atgctactcc tgttgtctgt actgcttatt caagtagtat tctaagttat gttactaaaa    660
aacatggtgg gtaaagcaca atcctaccca tcattgtcct ccaaaataat tgtatgacat    720
acacggccca gcccattgcc ctccctgcat ctctgtgctg ctttgccatt tcccctttcta    780
cccagcctcc tcaaggggta ccttggtgga tatttcagta cttaaaacca gactgtaatc    840
ataacctccc tctgtgtggc atcaataaat agccaaactc aaaaaaaaaa aaaaaaaaaa    900
aaaaaaaaaa aaaaaatatc ggtcgcaagc ttattccctt tagtgagggt taattttagc    960
ttgcactgcc ta                                                        972
```

<210> 18
<211> 1544
<212> DNA
<213> HOMO SAPIENS

<400> 18
```
tactttgact ttggatcatt tccctgactg ggctaatgtg acacatattg agacttagga     60
agagccacaa gaccacacac acagccctta ccctcctcag gactaccgaa ccttctggca    120
caccttgtac agagttttgg ggttcacacc ccaaaatgac ccaacgatgt ccacacacca    180
ccaaaaccca gccaatgggc cacctcttcc tccaagccca gatgcagaga tggacatggg    240
cagctggagg gtaggctcag aaatgaaggg aacccctcag tgggctgctg gacccatctt    300
tcccaagcct tgccattatc tctgtgaggg aggccaggta gccgagggat caggatgcag    360
gctgctgtac ccgctctgcc tcaagcatcc cccacacagg gctctggttt tcactcgctt    420
cgtcctagat agtttaaatg ggaatcagat ccctggttg agagctaaga caaccaccta    480
ccagtgccca tgtcccttcc agctcacctt gagcagcctc agatcatctc tgtcactctg    540
gaagggacac cccagccagg gacggaatgc ctggtcttga gcaacctccc actgctggag    600
tgcgagtggg aatcagagcc tcctgaagcc tctgggaact cctcctgtgg ccaccaccaa    660
aggatgagga atctgagttg ccaacttcag gacgcacacct ggcttgccac ccacagtgca    720
```

```
ccacaggcca acctacgccc ttcatcactt ggttctgttt taatcgactg gcccctgtc     780
ccacctctcc agtgagcctc cttcaactcc ttggtcccct gttgtctggg tcaacatttg     840
ccgagacgcc ttggctggca ccctctgggg tcccccttt  ctcccaggca ggtcatcttt     900
tctgggagat gcttcccctg ccatccccaa atagctagga tcacactcca agtatgggca     960
gtgatggcgc tctggggggcc acagtgggct atctaggtcc tccctcacct gaggcccaga    1020
gtggacacag ctgttaattt ccactggcta tgccacttca gagtctttca tgccagcgtt    1080
tgagctcctc tgggtaaaat cttccctttg ttgactggcc ttcacagcca tggctggtga    1140
caacagagga tcgttgagat tgagcagcgc ttggtgatct ctcagcaaac aacccctgcc    1200
cgtgggccaa tctacttgaa gttactcgga caaagacccc aaagtggggc aacaactcca    1260
gagaggctgt gggaatcttc agaagccccc ctgtaagaga cagacatgag agacaagcat    1320
cttctttccc ccgcaagtcc attttatttc cttcttgtgc tgctctggaa gagaggcagt    1380
agcaaagaga tgagctcctg gatggcattt tccagggcag gagaaagtat gagagcctca    1440
ggaaacccca tcaaggaccg agtatgtgtc tggttccttg ggtgggacga ttcctgacca    1500
cactgtccag ctcttgctct cattaaatgc tctgtctccc gcgg                    1544
```

```
<210> 19
<211> 1109
<212> DNA
<213> HOMO SAPIENS
<222> 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764,
765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779,
780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794,
795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 884, 885
```

```
<400> 19
cggacgcgtg ggggccagcc tggaggccca gacgtggcgc agcgactcgg aggttcgcct      60
ccagcttgcg catcatctgc ggccgggtcc cgatgagcct cctgttgcct ccgctggcgc     120
tgctgctgct tctcgcggcg cttgtggccc cagccacagc cgccactgcc taccggccgg     180
actggaaccg tctgagcggc ctaacccgcg cccgggtaga gacctgcggg ggatgacagc     240
tgaaccgcct aaaggaggtg aaggctttcg tcacgcagga cattccattc tatcacaacc     300
tggtgatgaa acacctccct ggggccgacc ctgagctcgt gctgctgggc cgccgctacg     360
aggaactaga gcgcatccca ctcagtgaaa tgacccgcga agagatcaat gcgctagtgc     420
aggagctcgg cttctaccgc aaggcggcgc ccgacgcgca ggtgcccccc gagtacgtgt     480
gggcgcccgc gaagccccca gaggaaactt cggaccacgc tgacctgtag gtccggggggc     540
gcggcggagc tgggacctac ctgcctgagt cctggagaca gaatgaagcg ctcagcatcc     600
cgggaatact tctcttgctg agagccgatg cccgtccccg ggccagcagg gatggggttg     660
gggaggttct cccaacccca cttctcttcct tccccagctc cactaaattc cctcctgcct     720
taaaaaaaac aagaaaaacc aaacaaacaa nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     780
nnnnnnnnnn nnnnnnnnnn nnnnntctt  ctatagtgtc acctaaattc aattcactgg     840
ccgtcgtttt acaacgtcgt gactggcgac ggacaaagtt atcnntttaa tcgccttgca     900
gcacataccc ctttggccag ctgggtaat  aggggaagcg ggccggaccc gatcggcctt     960
cccaaacagt tggggaagct tgaaatgcgc gacattgggc cgacggcctt ctatacggga    1020
ggatctctaa acgcggccgg ggtgttggtt gggttaaggc ggagtgtgac cccgcataat    1080
aacttttgca caggggccct ataggggcc                                     1109
```

```
<210> 20
<211> 1740
<212> DNA
<213> HOMO SAPIENS
```

```
<400> 20
aagagaagtt accccgatga cttggtttgg aaggggttaa ggcaccagtg catcctcttc      60
taaagtgatt tatgatgatg tgtggagttt aaaaacttta ccccacccca aagaacagcc     120
ctctcactcc tcactgagtc cactctgaac gtgctaaaat gggaggagg  cggtgttttg     180
ctgatctgtt aaattcttag tgaagtttcc ttgattccaa gtggctgctg ttgtttgagt     240
ttggtttgga gcaaaactga ggtagtccta acatttctgg gactgaatcc aggcaagaga     300
aagaagaaaa agaagaagaa aaagaggagg aaaaaggtag ggagaaataa agggaggaga     360
gaagcacagt gaaagaaaaa aaaagtccct tttgcgacat cacattcctg tgttttccct     420
cagcctggaa aacatattaa tcccagtgct tttacgcccg gaaacaaaga gactaagcca     480
gactatgggg gaaagggaga taagaaggat cctggaactt taaagaggga aagagtgaga     540
ttcagaaatc gccaggactg gactttaagg gacgtcctgt gtcagcacaa gggactggca     600
cacacagaca cacgagaccg aggagaaact gcagacaaat ggagatacaa agacttagaa     660
ggacagctcc tttcacctca tcctacttgt ccagaaggta aaaagacaca gccagaaaga     720
```

```
aaaggcatcg gctcagctct cagatcagga caggctgtgg atctgtggcg gtactctgaa    780
agctggagct gcagcacacc ccttttgtat tgctcaccct cggtaaagag agagagggct    840
gggaggaaaa gtagttcatc taggaaactg tcctgggaac caaacttctg atttcttttg    900
caaccctctg cattccatct ctatgagcca ccattggatt acacaatgac atggagaatg    960
ggaccccgtt tcactatgct gttggccatg tggctagtgt gtggatcaga accccacccc   1020
catgccacta ttagaggcag ccacggagga cggaaagtgc ctttggtttc tccggacagc   1080
agtaggccag ctcggtttct gaggcacact gggaggtctc gcggaattga gagatccact   1140
ctggaggaac caaaccttca gcctctccag agaaggagga gtgtgccgt gttgagacta    1200
gctcgcccaa cagagccgcc agcccgctcg gacatcaatg gggccgccgt gagacctgag   1260
caaagaccag cagccagggg ctctccgcgt gagatgatca gagatgaggg gtcctcagct   1320
cggtcaagaa tgttgcgttt cccttacggg gtccagctct cccaacatcc ttgccagctt   1380
tgcagggaag aacagagtat gggtcatctc agcccctcat gcctcggaag gctactaccg   1440
cctcatgatg agcctgctga aggacgatgt gtactgtgag ctggcggaga ggcacatcca   1500
acagattgtg ctcttccacc aggcaggtga ggaaggaggc aaggtgagaa ggatcaccag   1560
cgagggccag atcctggagc agcccctggg accctagcct catccctaag ctgatgagct   1620
tcctgaagct ggagaagggc aagtttggca tggtgctgct gaagaagacg ctgcaggtgg   1680
aggagcgcta tccatatccc gttaggctgg aagccatgta cgaggtcatc gaccaaggcc   1740
```

<210> 21
<211> 4467
<212> DNA
<213> HOMO SAPIENS
<222> 971, 978, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304,
1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316,
1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328,
1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340,
1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352,
1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364,
1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376,
1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388,
1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400,
1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412,
1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424,
1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436,
1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448,
1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460,
1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472,
1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484,
1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496,
1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508,
1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520,
1521, 1522

<400> 21
```
gcgtcgcgct caccctgcgc gtgcccccgc ggcgcctggg cgtcttcctg gactacgagg     60
ccggagagct gtccttcttc aacgtgtccg acggctccca catcttcacc ttccacgaca    120
ccttctcggg cgcgctctgt gcgtacttca ggcccagggc ccacgacggc ggcgaacatc    180
cggatcccct gaccatctgc ccgctgccgg ttagagggac gcgcgtcccc gaagagaacg    240
acagtgacac ctggctacag ccctatgagc ccgcggatcc cgccctggac tggtggtgag    300
gcgccctcgt ggccgcggga ctggcccogg gggggccccc tggatcccag ccagcgctt    360
tgctctcctg ctccgtctga agggagcagg tgcaccagcc aaaatgtcag cgaggggggac    420
aaagagaggg acctttgcct acgtagatgt gtatgtgtag tgcgattttc ttcaaggaaa    480
ggagacaagt ccaaagctcg tttgtggatt gtgggactga gcaaaggagt acaaatatat    540
ccacgtcgct cagagctggg gtgctcacgg tgggtggtgg gaaagaagcc agcatggaag    600
aaagaaggga gaaaactttg gtgactgcct tagagggatc agttaatttg tatagttta     660
tattttttgt atatgtttgc tagctctaaa aaggtcgaga tgcaataaca cttcgtaagc    720
aacgagttca cctaagtaag gctcagatcc tagtttttaaa aaccatttcc cattaaaatg    780
aagttggagg aacagctgct tctggagccg gggcaaaaaa tttcaaggtg agcctggagc    840
attgtgtgtg gtgaagtaaa ataaaggctc aaaacgtgac ggcaacccgg caaaagggta    900
gggagccagg ccgaagggcc tcactgacca attgtgggac aatttgaaca tcaggatgaa    960
taatgacagg ngaggttnta acacactgaa taaaaacata atccatgagt tcatgctgat   1020
actcaaattt cttttttaaaa aggagaaaca ggaaggtttc ttttggaggt gaaatctaat   1080
tattggtgag agtcttggag aacaggctgt ttccagtctc aaagcagtaa ccttatacac   1140
```

```
tacttataag tttgaaaggg gaaaggttac ctttacaatg gagacatcta ccagatgcat   1200
ccaagtgatt aaatttaaca tcatcaatga tgggaccaag gacattatta gtttgacaac   1260
tggggaaaga agtgttcttc accccctacc cccannnnnn nnnnnnnnnn nnnnnnnnnn   1320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1500
nnnnnnnnnn nnnnnnnnnn nnttcatttt cagagtgaga catttgtact gtggctatgt   1560
aggagaacat tcttgttctt agcaaacata ctgaagtttt tagatattaa ttaccacagt   1620
gtctgccact gaatttccag tgactaagtg gaaaaatata aaacatatga atataaagaa   1680
agaaagagac aagtcaaatg tagtaaaatg acaacacttg gtgactctag gtgactggtc   1740
gacagatgtt cattgtacta tcaatgtggc tttgctgtgg gtttgaaatt ttgcaaacta   1800
agagttgggt ggcggggaga aggatacacc aaaaaactaa gtgattatct ttggatggga   1860
aaatgtttgg taattgcatt cttaaaatgt cttctttgta tttttttaatg ttcaataatg   1920
tatatgtatc agttctgtaa taaaggggaa aacactttt ttaaatactc ataaaaaacc   1980
atccgtagga tcgagaagat caggcagaag ggctttgtcc agaaatgtaa ggcctctggt   2040
gtagagggcc aggtggtggc ggaggggaat gacggtggag ggggagcagg aaggccaagc   2100
ctgggcagcg agaagaagaa agaggaccca aggagagcac aagtcccacc aaccagagag   2160
agtcgggtga aggtcctgag aaaactggcc gccactgcac cagctttgcc ccaacctccc   2220
tcaacccccca gagccaccac ccttcctcct gccccaggcc acaacagtga ctcggtccac   2280
gtcccgggcg gtaacagttg ctgcaagacc tatgaccacc actgcctttc ccaccacggc   2340
agaggccctg gacccctca ccctcccaca ggcccctac aaccactgag gtgatcactg   2400
ccaggagacc ctcagtttca gagaatcttt accctccatc ccggaaggat cagcacaggg   2460
agaggccaca gacaaccagg aggcccagca aggccaccag cttggagagc ttcacaaatg   2520
ccctcccac caccatctca gaacccagca caagggctgc tggcccaggc cgttccgggg   2580
acaaccgcat ggacaggcgg gaacatggcc accgagaccc aaatgtggtg ccaggtcctc   2640
ccaagccagc aaaggagaaa cctcccaaaa agaaggccca ggacaaaatt cttagtaatg   2700
agtatgagga gaagtatgac ctcagccggc ctactgcctc tcagctggag gacgagctgc   2760
aggtggggaa tgttcccctt aaaaaagcaa aggagtctaa aaagcatgaa aagcttgaga   2820
aaccagagaa ggagaagaaa aaaaagatga agaatgagaa cgcagacaag ttacttaaga   2880
gtgaaaagca aatgaagaag tctgagaaaa agagcaagca agagaaagag aagagcaaga   2940
agaaaaaagg aggtaaaaca gaacaggatg gctatcagaa acccaccaac aaacacttca   3000
cgcagagtcc caagaagtca gtggccgacc tgctggggtc ctttgaaggc aaacgaagac   3060
tccttctgat cactgctccc aaggctgaga acaatatgta tgtgcaacaa cgtgatgaat   3120
atctggaaag tttctgcaag atggctacca ggaaaatctc tgtgatcacc atcttcggcc   3180
ctgtcaacaa cagcaccatg aaaatcgacc actttcagct agataatgag aagcccatgc   3240
gagtggtgga tgatgaagac ttggtagacc aagcgtctca tcagcgagct gaggaaagag   3300
tacggaatga cctacaatga cttcttcatg gtgctaacag atgtggatct gagagtcaag   3360
caatactatg aggtaccaat aacaatgaag tctgtgtttg atctgatcga tactttccag   3420
tcccgaatca aagatatgga gaagcagaag aaggagggca ttgtttgcaa agaggacaaa   3480
aagcagtccc tggagaactt cctatccagg ttccggtgga ggaggaggtt gctggtgatc   3540
tctgctccta acgatgaaga ctgggcctat tcacagcagc tctctgccct cagtggtcag   3600
gcgtgcaatt ttggtctgcg ccacataacc attctgaagc ttttaggcgt tggagaggaa   3660
gttggggggag tgttagaact gttcccaatt aatgggagct ctgttgttga gcgagaagac   3720
gtaccagccc atttggtgaa agacattcgt aactatttc aagtgagccc ggagtacttc   3780
tccatgcttc tagtcggaaa agacggaaat gtcaaatcct ggtatccttc cccaatgtgg   3840
tccatggtga ttgtgtacga tttaattgat tcgatgcaac ttcggagaca ggaaatggcg   3900
attcagcagt cactggggat gcgctgccca gaagatgagt atgcaggcta tggttaccat   3960
agttaccacc aaggataccA ggatggttac caggatgact accgtcatca tgagagttat   4020
caccatggat acccttactg agcagaaata tgtaaccta gactcagcca gtttcctctg   4080
cagctgctaa aactacatgt ggccagctcc attcttccac actgcgtact acattcctg   4140
cctttttctt tcagtgtttt tctaagacta aataatagc aaacttcac ctattcatga   4200
gttattattg aaacctcaaa tcataaagac atttaaaaga attgttttc taactgtgagg   4260
ggctctagtg ctaaataata gtactgaaaa ttgatattat tttccttttc ttatatgaag   4320
gaccttattt ggcatataaa attttataaa atatgtattt aaagctttt cttattttt   4380
gtattaattg gtaagtgaaa actctgttaa agatcacacc acaatgtttt caagaaacat   4440
ctgaaaagat aaaacaaaga acaaata                                       4467
```

<210> 22
<211> 2965
<212> DNA
<213> HOMO SAPIENS
<222> 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483,
1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495,

```
1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507,
1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519,
1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531,
1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543,
1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555,
1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567,
1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579,
1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591,
1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603,
1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615,
1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627,
1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639,
1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651,
1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663,
1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675,
1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687,
1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699,
1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711,
1712, 1713, 1714, 1715, 1716, 1717, 1718, 1719, 1720, 1721, 1722, 1723,
1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735,
1736, 1737, 1738, 1739, 1740, 1741, 1742, 1743, 1744, 1745, 1746, 1747,
1748, 1749, 1750, 1751, 1752, 1753, 1754, 1755, 1756, 1757, 1758, 1759,
1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768, 1769, 1770, 1771,
1772, 1773, 1774, 1775, 1776, 1777, 2948, 2951, 2961
```

```
<400> 22
  aaacaaagtt caatttagct ggatttctga actatggttt tgaatgttta aagaagaatg      60
  atgggtacag ttaggaaagt ttttttctta cacccgtgac ttgagggaaa cattgcttgt     120
  ctttgagaaa ttgactgaca tactggaaga gaacaccatt ttatctcagg ttagtgaaga     180
  atcagtgcag gtccctgact cttattttcc cagaggccat ggagctgaga ttgagactag     240
  ccttgtggtt ttcacactaa agagtttcct tgttatgggc aacatgcatg acctaatgtc     300
  ttgcaaaatc caatagaagt attgcagctt ccttctctgg ctcaagggct gagttaagtg     360
  aaaggaaaaa cagcacaatg gtgaccactg ataaaggctt tattaggtat atctgaggaa     420
  gtgggtcaca tgaaatgtaa aaagggaatg aggtttttgt tgtttttttgg aagtaaaggc     480
  aaacataaat attaccatga tgaattctag tgaaatgacc ccttgacttt gcttttctta     540
  atacagatat ttactgagag gaactatttt tataacacaa gaaaaattta caattgatta     600
  aaagtatcca tgtcttggat acatacgtat ctatagagct ggcatgtaat tcttcctcta     660
  taaagaatag gtataggaaa gactgaataa aaatggaggg atatcccctt ggatttcact     720
  tgcattgtgc aataagcaaa gaagggttga taaaagttct tgatcaaaaa gttcaaagaa     780
  accagaattt tagacagcaa gctaaataaa tattgtaaaa ttgcactata ttaggttaag     840
  tattatttag gtattataat atgctttgta aattttatat tccaaatatt gctcaatatt     900
  tttcatctat taaattaatt tctagtataa ataagtagct tctatatctg tcttagtcta     960
  ttataattgt aaggagtaaa attaaatgaa tagtctgcag gtataaattt gaacaatgca    1020
  tagatgatcg aaaattacgg aaaatcatag ggcagagagg tgtgaagatt catcattatg    1080
  tgaaatttgg atctttctca aatccttgct gaaatttagg atggttctca ctgtttttct    1140
  gtgctgatag taccctttcc aaggtgacct tcaggggat taaccttcct agctcaagca    1200
  aggagctaaa aggagcctta tgcatgatct tcccacatat caaaataact aaaaggcact    1260
  gagtttggca tttttctgcc tgctctgcta agaccttttt tttttttttac tttcattata    1320
  acatattata catgacatta tacaaaaatg attaaaatat attaaaacaa catcaacaat    1380
  ccaggatatt tttctataaa actttttaaa aataattgta tctatatatt caattttaca    1440
  tccttttttca aaggctttgt ttttctaaag gcnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1500
  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1560
  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1620
  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1680
  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1740
  nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnatc tgggccttac gtaatatatt    1800
  ttcttaatgg ctgcataata tcacatcaaa taggcatttt tcaaacctct ttccttatta    1860
  aacatgtaga ctatatccat tttttactaa aataaataac atttcagata atatctttgc    1920
  actgataatg ttgccaagcc atttctaaag tgaccttatc aatttaatta ccattggatg    1980
  agggtgttgc tttcatcgca ccattgtaga ttgtcttttt tatttcaatt tgcgtttatt    2040
  tataactggt tgcaaaggta cacagaacac acgctccttc aacttatctt tgataaaccc    2100
  aagcaaggat acaaaaagtt ggacgacatt gagtagagtc atggtatacg gtgctgaccc    2160
  tacagtatca gtggaaaaga taaggaaaat gtcactactc acctatgtta tgcaaaacag    2220
```

```
ttaggtgtgc tggggctgga tactgctctt ttacttgagc attggttgat taaagtttag   2280
gtaccatcca gggctggtct agagaagtct ttggagttaa ccatgctctt tttgttaaag   2340
aagagagtaa tgtgtttatc ctggctcata gtccgtcacc gaaaatagaa aatgccatcc   2400
ataggtaaaa tgctgaccta tagaaaaaaa tgaactctac ttttatagcc tagtaaaaat   2460
gctctacctg agtagttaaa agcaattcat gaagcctgaa gctaaagagc actctgatgg   2520
ttttggcata atagctgcat ttccagacct gacctttggc cccaaccaca agtgctccaa   2580
gccccaccag ctgaccaaag aaagcccaag ttctccttct gtccttccca caacctccct   2640
gctcccaaaa ctatgaaatt aatttgacca tattaacaca gctgactcct ccagtttact   2700
taaggtagaa agaatgagtt tacaacagat gaaaataagt gctttgggcg aactgtattc   2760
cttttaacag atccaaacta ttttacattt aaaaaaaaag ttaaactaaa cttctttact   2820
gctgatatgt ttcctgtatt ctagaaaaat ttttacactt tcacattatt tttgtacact   2880
ttccccatgt taagggatga tggcttttat aaatgtgtat tcattaaatg ttactttaaa   2940
aataaaanaa naaaaaaaaa naaaa                                          2965
```

```
<210> 23
<211> 1734
<212> DNA
<213> HOMO SAPIENS
<222> 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584,
585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599,
600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614,
615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629,
630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644,
645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659,
660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674,
675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689,
690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704,
705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719,
720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734,
735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749,
750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764,
765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779,
780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794,
795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809,
810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824,
825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839,
840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854,
855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869,
870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884,
885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898,
899
```

```
<400> 23
cgcctccgga aactgccccc cgggctgctg gccaacttca ccctcctgcg cacccttgac     60
cttggggaga accagttgga gaccttgcca cctgacctcc tgaggggtcc gctgcaatta    120
gaacggctac atctagaagg caacaaattg caagtactgg gaaaagatct cctcttgccg    180
cagccggacc tgcgctacct cttcctgaac ggcaacaagc tggccagggt ggcagccggt    240
gccttccagg gcctgcggca gctggacatg ctggacctct ccaataactc actggccagc    300
gtgcccgagg ggctctggac atccctaggg cagccaaact gggacatgcg ggatggcttc    360
gacatctccg gcaacccctg gatctgtgac cagaacctga gcgacctcta tcgttggctt    420
caggcccaaa aagacaagat gttttcccag aatgacacgc gctgtgctgg gcctgaagcc    480
gtgaagggcc agacgctcct gggcagtggc caagtcccag tgagaccagg ggcttgggtt    540
gagggtgggg ggtctggtag aacactgcaa nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnna    900
taatcctgct tttacaggtg aaactcgggg ctgtccatag cggctgggac cccgtttcat    960
ccatccatgc ttcctagaac acacgatggg ctttccttac ccatgcccaa ggtgtgccct   1020
ccgtctggaa tgccgttccc tgtttcccag atctcttgaa ctctgggttc tcccagcccc   1080
ttgtccttcc ttccagctga gccctggcca cactggggct gcctttctct gactctgtct   1140
tccccaagtc agggggctct ctgagtgcag ggtctgatgc tgagtcccac ttagcttggg   1200
```

```
    gtcagaacca aggggtttaa taaataaccc ttgaaaactg gatcggatga attggctttc    1260
    attgtgttcc tagcatcttc tcaaatcaac ttcccaggac tccagggtga aggaggaaaa    1320
    gaggcatggc ccaggccctg gggtgtggga tatggtctcc ctaggggatg acagttggga    1380
    tcaatggcct gtgacttctc ctctcccttc ccccatcctg ggacctaact ggaaataaaa    1440
    ccttgactgt tgcccgggtg tcattttacc agtggatttc tgccagggct tgtgtcctag    1500
    gagaaggttt aagttaaacc agattgccca ggtctccaaa cgatttgtca tgctgacctg    1560
    agatcatcga agggggcacc tgcccccggg caaggttgca ggggcaggat ggggctgaag    1620
    ggatgagcag ggtcccgggc ccacctgctg atacagcatt ggccatgtgg gggctgcaat    1680
    cggatttgga agaccctggg gcttgggggc atgtccattt tcccagtccc taaa          1734
```

<210> 24
<211> 4005
<212> DNA
<213> HOMO SAPIENS

<400> 24
```
    ggacaccgtc tgcagtggag tcactggtgc cgtaaatgtg gccaaggggg ccgtccagac      60
    gggctgtaga cacggccaag accgtgctga ccggcaccaa ggacacagtc actactgggc     120
    tcatggggggc agtgaatgtc gccaaaggga ccgtccagac cagtgtggac accaccaaga     180
    ctgtcctaac tggtaccaag gacaccgtct gcagtggggt gaccggtgct gcgaatgtgg     240
    ccaagggggc cgtccagggg ggcctggaca ctacaaagtc tgtcctgact ggcactaaag     300
    acaccgtatc cactgggctc acaggggctg tgaacttggc caaagggact gtccagaccg     360
    gcgtggacac cagcaagact gtcctgaccg gtaccaagga caccgtctgc agtggagtca     420
    ctggtgccgt aaatgtggcc aaaggcaccg tccagacagg tgtggacaca gccaagacgg     480
    tgctgagtgg cgctaaggat gcagtgacta ctggagtcac gggggcagtg aatgtggcca     540
    aggaaccgt gcagaccggc gtggacgcct ccaaggctgt gcttatgggt accaaggaca     600
    ctgtcttcag tggggttacc ggtgccatga gcatggccaa aggggccgtc caggggggcc     660
    tggacaccac caagacagtg ctgaccggaa ccaaagacgc agtgtccgct gggctcatgg     720
    ggtcagggaa cgtggcgaca ggggccaccc acactggcct cagcaccttc cagaactggt     780
    tacctagtac ccccgccacc tcctggggtg gactcaccag ttccaggacc acagacaatg     840
    gtggggagca gactgccctg agcccccaag aggcccgtt ctctggcatc tccacgcccc     900
    cggatgtgct cagtgtaggc ccggagcctg cctgggaagc cgcagccact accaagggcc     960
    ttgcgactga cgtggcgacg ttcacccaag gggccgcccc aggcagggag gacacggggc    1020
    ttttggccac cacacacggc cccgaagaag ccccacgctt ggcaatgctg cagaatgagt    1080
    tggaggggct ggggggacatc ttccaccca tgaatgcgga ggagcaagct cagctggctg    1140
    cctcccagcc cgggccaaag gtgctgtcgg cggaacaggg gagctacttc gttcgtttag    1200
    gtgacctggg tcccagcttc cgccagcggg catttgaaca cgcggtgagc cacctgcagc    1260
    acggccagtt ccaagccagg gacactctgg cccagctcca ggactgcttc aggctgattg    1320
    aaaaggccca gcaggctcca gaagggcagc cacgtctgga ccagggctca ggtgccagtg    1380
    cggaggacgc tgctgtccag gaggagcggg atgccggggt tctgtccagg gtctgcggcc    1440
    ttctccggca gctgcacacg gcctacagtg gcctggtctc cagcctccag ggcctgcccg    1500
    ccgagctcca gcagccagtg gggcgggcgc ggcacagcct ctgtgagctc tatggcatcg    1560
    tggcctcagc tggctctgta gaggagctgc ccgcagagcg gctggtgcag agccgcgagg    1620
    gtgtgcacca ggcttggcag gggttagagc agctgctgga gggcctacag cacaatcccc    1680
    cgctcagctg gctggtaggg cccttcgcct tgcccgctgg cgggcagtag ctgtaggagc    1740
    ctgcaggccc ggcgcggggt cgccctgctc tgtccaggga ggagctgcct cagaactttc    1800
    tccccgcccc caaacctgga tcggttccct aaagccctag acctttgggg ctgcagctgg    1860
    ctgagcgccg aggggctgcg gaggcagtga ccttcttaac tgagccaccc cacgcccctgc    1920
    tccgggcctg cctgacatctc ccacctcctc cccagcgctg cctgcccctc tcggagcctg    1980
    gggtcactca gaccaccagc caagagcctt cccttgaagt ccccaagcaa gcactgcaat    2040
    taggaaagag aaaaagcagc gtgcccagcc tggaagggca tctgtttgcc ccgctagcaa    2100
    cccttttata tctagcaggg ctcttccagt cctgcagcac gggcccccag ctatcagcgg    2160
    tgcaggcagt gctgtggcat cccaggctcc gggcagctcc gttctcatgc tgaaagtggg    2220
    tctccggcct tagcacacac accttgaggg tcttaagaac cacattccct catagtagaa    2280
    agtactagaa aaagcgacac tgccatcatc atcccaaggc aggctgctac tgcctttgct    2340
    gacccccggg gtggcctcac ggtggggaca aagctgccag gagccacagc agccacagct    2400
    ggggctttgc accagcctgg cttgagactg agcagtttgc aggggggtggg gggtgcaaaa    2460
    aacaagcaaa caggctgctg ctgcctccag ctgcccacca caggcctgcc ccaggcacct    2520
    ggggctctga ggcccctggg gaggctgggc ccagcagctg cccctggaga acacagacaa    2580
    aggacttccc cgcagggaac tgtgccctat ggagggatca gacagggctg ggaacagcca    2640
    cagaggctgc gtgcctatgg cacagccctt cctccgccgc acactccccc tgggtcctca    2700
    ggcccaccca agcgccgggc tgcagaggaa gcggggctgg ggaggctgca ggcatcagag    2760
    acactggtgg tggcggaccc ggccgccggg ccccgtgctc tcaggctagc ccaggtcgtg    2820
```

```
gaggctggca ggctcaggtc gggtgtgaga cgtgccgtgg ctgcgctcag tccagcgggg   2880
aggagccgtt cagcccggcc tccccaggaa gccatatccc cactcacccg gtaagagaac   2940
cttgtcgtcc cctttccatg ctctcctagg acacgagccc aggaacccca gacccagggg   3000
gaggaagggt ggaggggccc caggggtcac catgtgcacc aggggccgtg aggggccggg   3060
gcattcagct cagctctgaa ccggggaagc tggcacggca aggactgcct caggtgacgg   3120
gccgtgagag gggacgggtc aggagccttc ccaagccttc tcctcagcgc gacacccatg   3180
gccatcggag gctaggatgc cagacacagc catttgcaga aatcaggcac agtgactgca   3240
gctcacgtcc agccaaccaa gcatggggcc gcagctcagg aagtcccttc ccgccacacc   3300
acagcctaat tcttactggg acggaggcaa ctcggctacg ctgggcagga cgacaaacac   3360
gagacgccac tgtggaatga gcaacttcgg agcacggggt gacttgcttg ggaccgtgcc   3420
cacgtgacag ccccttatgc agaggaggaa agagaagccc cgagtgggag gggaacctgt   3480
ccaaagtcac acggtgtgtg ggtgacacag ctggggtgag tcgaggctgg cccctgaggc   3540
ccatgctccc tgaacgctgg agaccactgt cggctagcag cggctctcag ggaaggcctg   3600
gtctccaccc tcccagccta gcctcgcgga ccctcgtcct ccccacatcg gacctgctca   3660
cctgcctgga ccctgggctg ccagatgcag gaagcatcaa accccccagc ctcgtgggtg   3720
cggggcaggg cgcaggcagc acagcttaga tgccctggtt tgtccctctt gtctcctggg   3780
aagagcttgc tcccgcccag ctctcctgcc actggccttt cagggttggg ctgggcccag   3840
agtgcctttt agtcgcttct cacggtggcc tgatggctca acccagtccc aaacgggccc   3900
agtgacactg ccgactgcac cccagctcag gcccccactg caccagcaat gctagaaaac   3960
caagccaata aaagtgattt ctttttttcat taaaaaaaaa aaaaa                  4005
```

<210> 25
<211> 846
<212> DNA
<213> HOMO SAPIENS

<400> 25
```
caaaaatgag cggggtgtgg tggcccatgc ctgtagtccc agctgctcgg gagactgaat     60
ctcttgagcc tgggaagcag aggttgcagt gaactgagat cgcgtcactg cactccagcc    120
tggtgacaga gcgagattcc atctcaaaaa aaaaaaacag tatgcacgta caaatttctt    180
aacctgttat caatgtctga gctacataat tatctttcta gttggagttt gttttaggtg    240
tgtaccaact gacatttcag ttttttctgtt tgaagtccaa tgtattagtg actctgtggc    300
tgctctcttc acctgccct tgtggcctgt ctacaattct aaatggattt tgaactcaat    360
gtcgtcgctt ctggtttcct gcatatacca atagcattac ctatgacttt ttttttcctg    420
agctattttc actgagctga gctaatgaac taaaactgag ttatgtttaa tatttgtatc    480
aaatacataa aaggaatact gcttttcct tttgtggctc aaaggtagct gcattttaaa    540
atatttgtga aaataaaaac ttttgttatt agaaaaaaaa aaaaaaaaaa aaaaaaaaaa    600
aaaaaaaaaa agaaagacca aaaaggaaga gaaaggaaaa agaagaagag aaacggaggg    660
acaacgggaa acacagagag cgagccggtg acgaaaagcg ggaaggccaa cgcaggagaa    720
gaaagagagg ggggcggcgt cgctcattgt gggagtgtcc tcagagttat gcgagtgggg    780
gatgatgggc aggagtgcta tgcgcccctt tgtatgaggg ggtgcctcaa ttgttgatgg    840
gccggg                                                              846
```

<210> 26
<211> 599
<212> DNA
<213> HOMO SAPIENS
<222> 103

<400> 26
```
cggacggtgg gcggacgcgt gggcggacgc gtggggggtc ttgctatgtt taacaaccca     60
ggctgatctc aaactcctgg gctcaaatga tcctccacct tgnccctca aagtgatggg    120
attataggca tgagccactg gctggcatca ggtgccaaga tttctgtact gcctctaatt    180
tctgctacca cttaaactca ggcaggtgga gcctacacac tgatatttcc ttgtggatat    240
cacacttcag aacgtgtccg ctagataaag ctctcaaact taccaaggaa agtgatgaca    300
gcttgactcg gccttacaca gaaccctatg taggtctcac acaatagaac aatgtacaaa    360
taagcatttt tctttcccaa agaagcatgt aaagatttcc cattcctgcc actcaacttc    420
tctttgttgt gacagggtgg aagaattact gtatatagaa aagatgtccg cagcgttcag    480
taaacacaga cactaatgag actcagaggc tcatctgtgg tcaggtatta taacagctta    540
aaactaaaaa aaaaaaaaaa aaaagggcgg tccaagctta ttcccttag tgaggttaa    599
```
```

```
<210> 27
<211> 603
<212> DNA
<213> HOMO SAPIENS

<400> 27
 gttccacgtt gcttgaaatt gaaaatcaag ataaaaatgt tcacaattaa gctccttctt        60
 tttattgttc ctctagttat ttcctccaga attgatcaag acaattcatc atttgattct       120
 ctatctccag agccaaaatc aagatttgct atgttagacg atgtaaaaat tttagccaat       180
 ggcctccttc agttgggaca tggtcttaaa gactttgtcc ataagacgaa gggccaaatt       240
 aatgacatat ttcaaaaact caacatattt gatcagtctt tttatgatct atcgctgcaa       300
 accagtgaaa tcaaagaaga agaaaaggaa ctgagaagaa ctacatataa actacaagtc       360
 aaaaatgaag aggtaaagaa tatgtcactt gaactcaact caaaacttga aagcctccta       420
 gaagaaaaaa ttctacttca acaaaaagtg aaatatttag aagagcaact aactaactta       480
 attcaaaatc aacctgaaac tccagaacac ccagaagtaa cttcacttaa agtaagtaga       540
 aaataaagag ggttcatgtt tatgttttca atgtggatct tttaaaaaaa atatttctaa       600
 ggc                                                                     603

<210> 28
<211> 879
<212> DNA
<213> HOMO SAPIENS

<400> 28
 gccacgcgtc cgcaaacaca aaaagaaaac aaacaaacaa aaaaacaaaa agactggctg        60
 gcgggaaggg tgactcgggc ctttgctccc gagccagagc ccccaacccct gacctgatcc      120
 ccctctctgc gcaggtggag ttctacttcc tttcccagta cgtgtcgcca gccgactccc       180
 cgttccgcca catcttcatg ggccgtggag accacacgct gggcgccctg ctggaccacc       240
 tgcggctgct gcgctccaac agctccggga ccccccgggggc cacctcctcc actggcttcc     300
 aggagagccg tttccggcgt cagctagccc tgctcacctg gacgctgcaa ggggcagcca       360
 atgcgcttag cggggatgtc tggaacattg ataacaactt ctgaggccct ggggatcctc       420
 acatccccgt cccccagtca agagctcctc tgctcctcgc ttgaatgatt cagggtcagg       480
 gaggtggctc agagtccacc tctcattgct gatcaatttc tcattacccc tacacatctc       540
 tccacggagc ccagacccca gcacagatat ccacacaccc cagccctgca gtgtagctga       600
 ccctaatgtg acggtcatac tgtcggttaa tcagagagta gcatcccttc aatcacagcc       660
 ccttccccttt tctggggtcc tccataccta gagaccactc tgggaggttt gctaggccct      720
 gggacctggc cagctctgtt agtgggagag atcgctggca ccatagcctt atggccaaca       780
 ggtggtctgt ggtgaaaggg gcgtggagtt tcaatatcaa taaaccacct gatatcaata       840
 aaaaaaaaaa aaaaaaattc tgggcgcaag aaatcgctg                              879

<210> 29
<211> 397
<212> DNA
<213> HOMO SAPIENS
<222> 319, 331

<400> 29
 cctactcaac aggggggtccc aaatgcccac tgcaactagg tacagggtct gtgtgtggtg        60
 gtggaggatt ggcattggga gacatgggag gcaaagagct gggcctggcc aggccaggcc       120
 tctggcttcc aagaactcct agttccaggg gacaccagt gggggaagat ctggctgctg         180
 ggaggcccac agcctagggc tggtcggcca aacagccagc tctggtccct gctcacaagt        240
 gccctatggc ttctaatgca tttctcttct tcctcccttg ctgcacctgc agatgcagaa       300
 ggcagggcct ccccagcanc ctactggctg nccactcca tttggactgg cacattggac        360
 tggggcatca cattccctca gaacagcctg ataaatg                                397

<210> 30
<211> 1740
<212> DNA
<213> HOMO SAPIENS
<222> 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229,
230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244,
245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259,
260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274,
```

275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289,
290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304,
305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319,
320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334,
335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349,
350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364,
365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379,
380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 846, 847, 848, 849,
850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864,
865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879,
880, 881, 882, 883, 884, 885, 886, 887, 888, 889

<400> 30
```
tacaaagttt taagaaagcc agcatctcag aaaggccttt caaacaagga cacttaatta        60
gccatcttat gtataagaaa agaaatataa agaacatgaa aatttaaaaa cagatttggc       120
agttttataa cagtctagga ggtggtgtta tttttttccta ttaagaatta gagggcaggt      180
taggaataaa taaaatacag tttgaaaata atgagnnnnn nnnnnnnnnn nnnnnnnnnn        240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn        300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn        360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gcacttcccc tactgattgc tgccttctct       420
gtggctacaa gggacccaca gaattacagg gaagttacag ggaagcaggt ttcatctcaa       480
tattggggaga gatttcaaac aatcacacct gcctgagaag gagtgggctg tcactaggaa      540
tttttattcc cagtccgtca ggaattttgt agaagggctt catgtgctgg taccaatagg       600
acaggaagat tttaatcagc tttactatct atgtttttt atggaaactg tgtgtatgta        660
tacatacatt ttccaaaaag aaaaattaaa tgattataga gattatgttt ttcagactac       720
tcacgtatct gcttttctta ctccccacct ctgctgataa ttcctagttc attggttttt      780
cccccacact ggaattacct ggggagctta aaaaaccctg atgcctgggt cccaccctca       840
gagatnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnt gatatgaatg        900
cagcctgggc actgggaaat ttaaaaactc cccagataat tactgtgaca gccaaggttg       960
agacctccta gtctagagct ttgctataca cagggtggtc cacaagccag gagcatcagc      1020
atcacttggc agcgcttaga gattcagacc ccagaccac tgaatctgac cctgcatttt      1080
cactagaccc caggtgatca gctgcactag actcaacctc taaaccaaga cctcccaccc     1140
tcacagtcta tgatccttta gtgaccctca gctgagtcct gtgctgaact gtgtttgttc     1200
tccttgagca catgcccgct gaccagggac agactggatg agcaagcaac ctgctggcta     1260
tggagaagag ccaggctggg taaatgtttg ctgtgactaa gccaggatca aagaactgcc     1320
tgttgcttgc actggctggc actgagcttg ccactctgtg aactgtgctt ccttccctg     1380
catggacctg tgcctcagtc actattatcc gcaggcctc tccaagggca gccctctcct     1440
tgtttatccc tcttaagcct gcgtgcagga aggcacatta accctgtggc ccctgcagg     1500
caggagggtg ttgggtgccc ttacctacct tgcccttttt cttgtaccgt aggctgtgcc     1560
gtttatgagt aagtgatgtg tgtctgtgtg tgtgtctaga agtgctgcac tcaccttgtg     1620
ttattggagg ttgtgtaacc ccctagcttt gagcctggtc tcagatgttc ctttcccgt      1680
tctctgtcca gccgttaacg cccccagtct gtaataaaag cctatcagcc gtgcactta       1740
```

<210> 31
<211> 2394
<212> DNA
<213> HOMO SAPIENS

<400> 31
```
aaatgtagaa gggaaagtcc ttcctggtag taatggaaaa ccgaatggac agagaattat        60
caatggcct caaggaacaa agtgggttgt ggaccttgat cgtgggttag tattgaatgc        120
agaaggaagg tacctccaag attcacatgg aaatcctctt cggattaaac taggaggaga       180
tggtcgaacc attgtagatc tggaagggac ccccgtggtg agtcctgacg gcctcccact       240
ctttgggcag gggcgacatg gcacacctct ggccaatgcc caggataagc caattttgag       300
tcttggagga aagccgctgg tgggcttgga ggtcatcaaa aaaaccaccc atcccctac        360
cactaccatg cagcccacca ctactacgac gccctgcct accactacaa ccccgaggcc        420
caccactgcc accacccgcc gcacgaccac caggcgtcca acaaccacag tccgaaccac        480
tacgcggaca accaccacca ccaccccaa acccaccact cccatcccca cctgtccccc         540
tgggaccttg gaacggcacg acgatgatgg caacctgata atgagctcca atgggatccc        600
agagtgctac gctgaagaag atgagttctc aggcttggag actgacactg cagtacctac       660
ggaagaggcc tacgttatat atgatgaaga ttatgaattt gagacgtcaa ggccaccaac        720
caccactgag ccttcgacca ctgctaccac accgagggtg atcccagagg aaggcgccat        780
cagttccttt cctgaagaag aatttgatct ggctggaagg aaacgatttg ttgctcctta       840
```

```
cgtgacgtac ctaaataaag acccatcagc cccgtgctct ctgactgatg cactggatca      900
cttccaagtg gacagcctcg atgaaatcat ccccaatgac ctgaagaaga gtgatctgcc      960
tccccagcat gctccccgca acatcaccgt ggtggccgtg gaaggttgcc actcatttgt     1020
cattgtggac tgggacaaag ccacccoagg agatgtggtc acaggttact tggtttacag     1080
tgcatcctat gaagacttca tcaggaacaa gtggtccact caagcttcat cagtaactca     1140
cttgcccatt gagaacctaa agcccaacac gaggtattat tttaaagtgc aagcacaaaa     1200
tcctcatggc tacggaccta tcagcccttc ggtctcattt gtcaccgaat cagataatcc     1260
tctgcttgtt gtgaggcccc caggcggtga gcctatctgg atcccattcg ctttcaaaca     1320
tgatcccagc tacacggact gccatggacg gcaatatgtg aagcgcacgt ggtatcgaaa     1380
gttcgtggga gttgttcttt gtaattcact gaggtataaa atctacctca gtgacaacct     1440
gaaagataca ttctacagca ttggagacag ctggggaaga ggtgaagacc attgccaatt     1500
tgtggattca caccttgatg gaagaacagg gcctcagtcc tatgtagaag ccctccctac     1560
tattcaaggc tactatcgcc agtatcgtca ggagcctgtc aggtttggga acatcggctt     1620
cggaaccccc tactactatg tgggctggta cgagtgtggg gtctccatcc ctggaaagtg     1680
gtaatcacag gaccgtcatg ctgcaagctt gccctgccca gccccaccaa ctaagtcgca     1740
ctaggggctg tgagcaaaga cagccagcat gctcagcccc gctgccctag gtgccaggaa     1800
ggtcacagat ggacactggc cattctggtc atctcagtct ggaactcagt cccacttctt     1860
ggcctggaca atgaacagga ttcagttttg ctgttaactt tgcttctcta cttttttttg     1920
tttgtttgta atagcacatc ccagagacat cagaaaccag caactgattc agtgtgattt     1980
ccagactttt taggcatgaa attcggacac ttcagtattt ccaggaatag catatgcacg     2040
ctgttcttgc ttcatggaat gctacatgct ttctgttttt ctcattttgg atttctccaa     2100
aactaactga atttaagctt caggtccctt tgtatgcagt agaaaggaat tattaaaaac     2160
accaccaaag aaaataaata tatcctactt gaaatttact ctatggactt acccactgct     2220
agaataaatg tatcaaatct tatttgtaaa ttctcaattt tgatatatat atgtatatat     2280
gcatatacat atccacactt gtctgcaaga atattgatta aaattgctaa atttgtactt     2340
gttcaccaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aagg            2394
```

```
<210> 32
<211> 499
<212> DNA
<213> HOMO SAPIENS
```

```
<400> 32
ctgggatagc aataacctgt gaaaatgctc ccccggctaa tttgtatcaa tgattatgaa       60
caacatgcta aatcagtact tccaaagtct aaatatgact attacaggtc tggggcaaat      120
gatgaagaaa ctttggctga taatattgca gcattttcca gatggaagct gtatccaagg      180
atgctccgga atgttgctga aacagatctg tcgacttctg ttttaggaca gagggtcagc      240
atgccaatat gtgtgggggc tacggccatg cagcgcatgg ctcatgtgga cggcgagctt      300
gccactgtga gagcctgtca gtccctggga acgggcatga tgttgagttc ctgggccacc      360
tcctcaattg aagaagtggc ggaagctggt cctgaggcac ttcgttggct gctactgtat      420
atctacaagg accgagaagt caccaagaag ctagtgcggc aggcagagaa gatgggctac      480
aaggccatat ttgtgacag                                                   499
```

```
<210> 33
<211> 1774
<212> DNA
<213> HOMO SAPIENS
<222> 1679, 1681, 1684, 1685, 1686, 1691, 1692, 1693, 1705, 1706, 1707,
1708, 1710, 1714, 1716, 1731, 1740, 1744
```

```
<400> 33
ctttcctaga caaggctgaa aggggccaac attatttctg aagacttcat tattggaatt       60
ctatgggagt gatctcactg agctattttg gaatagaaat gtggctagtt gcctgacctc      120
cctcaatggt ttcacgtggc tttcaaaggg aaggaagggc agtgctgact tttggtaaaa      180
tgggcgaaag ggtccatgcc agcaacacaa tcactcaaag tccagatgag ggatcagtaa      240
atacaacgtg cctgaaaggt ggcccttgag cacattcctc cggtagacat taacttatta      300
aattgattct gattacaaat ataaactttg cccccatctc acccagtaac aatgcaagag      360
ttgatgtcag tctataaaag gaagtaggaa ctgtccctgg ctttcaggct ccaacatcct      420
cccctgtca agatgtggca cctcaaactt tgtcgagtcc tcatgatctt cctgttgctg      480
ttggggccagg taaggaggga aggatactta tgtgtgtgtg tggagtgtgg agatgatagt      540
ggtggtggaa cttgaaagct agattcagtc ctgaggaatg gttcctctgt tctgagtcta      600
cagcatctgc ggaatggaat gatcactctt ccaaggtgtg cagcagggtg tcaacacttt      660
catatctgaa tgtctttgcc cttacagata gatggctccc caataccaga agtgagttcg      720
```

```
gcaaagagaa ggccacggag aatgacccca ttttggagag gggtttccct caggcctatt      780
ggagcctcct gccgggatga ttctgagtgt atcacaaggc tatgcagaaa aagacgctgt      840
tccttaagtg tggcccagga atgatgtaca taccagggaa agaaaggaca gcagtcacct      900
ccgacaatgc tccgttctat ggaatattga ttaactgcat tttggctgga gacacccaag      960
tgaagcaatc ttgtattttt aatatttaaa ggcagatgta cgctttaaat tggtctccat     1020
ttcttcttag aatgttgata tatggataag cataactaaa cttgtcaatt tagagtttat     1080
ttttctatgg atactattaa atgtctcaaa ttgaaatttt agcagtctgg aattcaagct     1140
tttgagggaa agaaggattc actttgtata ctaaagaaaa aaacagcatt gcccaataat     1200
gtgttaactt ctcaatctgg aaagtgtagt gagagctaca taatcaatag ctacgtaatc     1260
aacttcagca agttcctaag ctgtggccct ggatcccttc actccatact cttcagggag     1320
gtgtcaaagg tggtcaagct tgggaggctg aggcaggaga atggcgtgaa ccgggagacg     1380
gacttgcagt gagccgagat ccgccactga ctccagcctg gggaaagagc gagactccgt     1440
tcacaaaaaa aaagaatcaa aaaaaaaaag gggagccccc ccttggtatc ggaagaccca     1500
gtcctgtaat tcacacaggt tgagttcaag gcattaagcc ctgtaagggc cacttcggcc     1560
cctcagagtt gctgttctga tccaacggaa gccgcttaca aatttccctt cggaatttgc     1620
ctccggcatt ccctaggggc ggtatttgga agcaaagtcc ttttaacagc cagtgtatnc     1680
naannncggg nnngcccttg cggcnnnngn ccananattg ctccttcttc ncctcttctn     1740
tttnttcccc cccgtgtcga caggggggtgt ggtc                                1774
```

```
<210> 34
<211> 4158
<212> DNA
<213> HOMO SAPIENS
<222> 3667, 3668, 3669, 3670, 3671, 3672, 3673, 3674, 3675, 3676, 3677,
3678, 3679, 3680, 3681, 3682, 3683, 3684, 3685, 3686, 3687, 3688, 3689,
3690, 3691, 3692, 3693, 3694, 3695, 3696, 3697, 3698, 3699, 3700, 3701,
3702, 3703, 3704, 3705, 3706, 3707, 3708, 3709, 3710, 3711, 3712, 3713,
3714, 3715, 3716, 3717, 3718, 3719, 3720, 3721, 3722
```

```
<400> 34
ctcccacaac aatttcattg ttgttagcat atctatttct ccatacattg taaaactgta       60
atccttaggt atttctaaaa cataaagagg agaattaagt cagctgcaga acaatggggc      120
tgattcttct gctttttctc tggaaaatct ttcattgctt ttggtggaaa tttacctaga      180
ggttacaacc acaggatgta gcttggtctc ttatttgcct ttttgggaaa ccaattaaga      240
ttaatacagg ataaaggaaa aaagcaatct attcattata taacacagtt gtttgtatta      300
cttgttccct gcaaaggaaa tctgttgaat gcttgcattt tgaattcttt tctaatagaa      360
caaccaaaaa aggcttctta tggtgcagca ggaaaaaga tcattttat agctttgcat         420
tcttaacata gcatttaaag agcggcatga attagaggaa agacatggaa cacacaggta      480
gtcggtttga gatcatcggc ttaaaagtat cctaggatgg taatgaccca gaagtatttc      540
cagttgtcta gtggtgtggt atgcaggaat gagaagtgtt ttctttccat ttcctgttgg      600
acaggtggca atcttagcag agccactatt tggagttgat aactaaagat gcaaataacg      660
tgactatgcc ttctgtcat cctacgacta tttggagttc tccaaaacct tgtaagaggc       720
atgtcaggca tgcagtaaaa gcatctacaa cttcagctgg gcactggcag cataggtctc      780
atcttggacc atacagtccc actttataga agagagtgga agttctccaa aacaatatcc      840
acaacaaagt ctgacctcac tctgagggag atgggaagtg ggaggaagaa ggactaacca      900
gctccctgga gtaagaggaa tttgctttcc ctgtctgccc accagggget atatgtgcca      960
cctttcaggt tggggccaag gaagtgatgt cagtgtgaca gaagggagag ttagacctcc     1020
agacgtcagc ctccctccca tggggtacat tttcaatctg agtgttgttg ccttagctgt     1080
gttggtatta gcttgattgg ttggtccgct ggttatgagg tgtagggagg cagttttgt      1140
ttagttttta ggactttgcc tcttcctttg tccttagcat aatttctagg cagagcatcc     1200
acgaagtcgg ttttcattgc cagctcaaga gcgacaatca tttacgagtt cctatgttat     1260
gttaggtgcc ttatgtatat tatcccaaat ccactgcatg gtttaaatac aggcactgga     1320
atataaatga aaaaggtcat tacagtcact gactttctgc aggaccttaa acatttctct     1380
ttccacaagt ttccccttaa tcatgtgtca aacctctctt cctgacggga atgttgtgct     1440
ataatgaatc tgcataacgc ttgggattct aggaggaagg aaggttccat ggacatgtaa     1500
gtacagcata ttcccctcag tcttctagga gggcagagtg aatcccagaa ctggtaagat     1560
tgggaatctg agcattgcca cttaatcttt agaatattta tcattttgac acatcctgtt     1620
tttagagag gaaaacaaac acagtttctg cattggtagt gtaaagcata ccttgttagg      1680
aacgtgtttt gtaagacaca tttgggttgt cattctagag catgtcaaac tttgtacttc     1740
aaaatatatt tagtatgatt gttagtggta acatatatca aggctttgaa ttaactgttt     1800
tatttaattt tcacaagaag cacttatttt agccatagga aaaccaatct gagctacaaa     1860
tagttcttta aaataagccc aggttattta gctattctag aaagtgccga cttctttcaa     1920
gaagcaggca ttgtaggaca gctgagaatt atcacatagc ctaaattcta gcctggcagc     1980
```

```
aagagtcaca tctgagatgt ccaaaaaaaa aaaaaaaaaa cacctgatct acattgaaag    2040
ggggtagact aacgtatgtg agaccatttt cctatttgca gttacaaggt taaagaactt    2100
tgaaggtcat tcggctgcta agaggcatgt cgaacactct gtgtggctct ttcacagtaa    2160
accctcctaa gagcagaaga cacatggctg ttagtgtctg cgtttagatt taatttctca    2220
aataaaggcc cttggctgcg tatcatttca tccagttata aactagggct cctgcaagca    2280
cccccattct aagggtgaat tattgaaatc agttgctatt tgatgagtca caactggccc    2340
agcaggcagg gcatttgaag tcatggtcat caaaaagaaa tgattgtttt ttgaaaagct    2400
aaatgcttaa aatgcttcta gagggaagtc gtggggcgtg tgctcattct ctttaaaatc    2460
agggttgttg agtttgtttt taaacatttt tataagttca tgagaaaaaa tatataaatt    2520
ctaagaacca acactgtatt cccagaaaca tgaccctcgc tggtcttggg tccacatatc    2580
attggactct gggggacaca aagatgcctg tgacactttg gtgttgccga gttagtcaac    2640
aattattctg ggaaaaagca gaattgaatt cttctctaga tgtcctacca gggttggcca    2700
agggccacaa agcaggctaa taaattccca caggatccag acaccaggca aaattgctct    2760
aagaagccag ttactgtcat ccctctatgg ttctagaaaa aatagtacaa aaatgacagg    2820
tcatcctatg agcgtcatgc caatgaaacc ccatcttctg gagaagccct tgaatcagaa    2880
ttatcttttt tcttgatgtc gtcagatgca gccagtttct taatttttt aaaaactgta    2940
tgtttctgtg gtatgtatat ttgtacacct aactacctgg cacttggaaa tcacagcact    3000
actcagaggc aattgaataa agagaaattt aattttaaat atcaagtcct gtcaaacatt    3060
tctcaaactt ctgattttat caaaggtttg ccagccaata aagtgcatcc caagtataca    3120
ggggagaaag ctagactcct acagggtcct agagtttaag taatttttt gttattaata    3180
taggtaataa ttttctaat ttttattttt tggttccaaa tgtaaagctc cttgtgttta    3240
cctctgttta tgtcattctt gacatgttta tctaaattat gtgtgctctg tgacaggtga    3300
aatgtaaatc tgggatccat agtcaagata tcataaggac ctacttccca gcctaccttt    3360
cttcctctac ctgataatga taatactcaa aataacaaca ttcaaaggaa acacaaagaa    3420
atcctgcttt cacatctcct atttcttggg ctccttaata actactgatg gtttgttcat    3480
gaaaaaaaat tttaaatca aagattgta cttggccctg agttcaaaaa atttcaaaaa    3540
tcaaaagttt gtacttggcc ctgagttgaa aaaaaaatt cacattctaa gaataaacag    3600
aaaaatgttc ttcttggaag taaataacaa aaaaaaaaa aaaaaaaaa aaaaaaaaa    3660
aaaaaannnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn    3720
nntcttctat agtgtcacct aaattcaatt cactggccgt cgttttacaa cgtcgtgact    3780
gggaaaaccc tggcgttacc caacttaatc gccttgcagc acatcccct ttcgcaagct    3840
ggcgtaatac gcgaagaggc ccgaaccgtt ggcccttccc aacagttgcg cagcctgaat    3900
ggcgaatggg acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt ggttagtaga    3960
ggtgtgccgt aaaaaataga ataatttttt ttcaagagat gagcagaatt gagtaggaat    4020
gattacgggg aggaaaagat ctagaagata gacaatagag aggagagaaa aagagggacg    4080
aggaggctga gaggaaaaga gtagaagcgt gatatgaata tatacagaaa cagaaaaagg    4140
agagagggta agacataa                                                  4158
```

<210> 35
<211> 366
<212> PRT
<213> HOMO SAPIENS

<400> 35

```
Met Ala Leu Arg Phe Leu Leu Gly Phe Leu Leu Ala Gly Val Asp Leu
1               5                   10                  15
Gly Val Tyr Leu Met Arg Leu Glu Leu Cys Asp Pro Thr Gln Arg Leu
            20                  25                  30
Arg Val Ala Leu Ala Gly Glu Leu Val Gly Val Gly Gly His Phe Leu
        35                  40                  45
Phe Leu Gly Leu Ala Leu Val Ser Lys Asp Trp Arg Phe Leu Gln Arg
    50                  55                  60
Met Ile Thr Ala Pro Cys Ile Leu Phe Leu Phe Tyr Gly Trp Pro Gly
65                  70                  75                  80
Leu Phe Leu Glu Ser Ala Arg Trp Leu Ile Val Lys Arg Gln Ile Glu
                85                  90                  95
Glu Ala Gln Ser Val Leu Arg Ile Leu Ala Glu Arg Asn Arg Pro His
            100                 105                 110
Gly Gln Met Leu Gly Glu Glu Ala Gln Glu Ala Leu Gln Asp Leu Glu
        115                 120                 125
Asn Thr Cys Pro Leu Pro Ala Thr Ser Ser Phe Ser Phe Ala Ser Leu
    130                 135                 140
```

21/22

```
Leu Asn Tyr Arg Asn Ile Trp Lys Asn Leu Leu Ile Leu Gly Phe Thr
145                 150             155                 160
Asn Phe Ile Ala His Ala Ile Arg His Cys Tyr Gln Pro Val Gly Gly
            165                 170                 175
Gly Gly Ser Pro Ser Asp Phe Tyr Leu Cys Ser Leu Leu Ala Ser Gly
            180                 185                 190
Thr Ala Ala Leu Ala Cys Val Phe Leu Gly Val Thr Val Asp Arg Phe
        195             200                 205
Gly Arg Arg Gly Ile Leu Leu Leu Ser Met Thr Leu Thr Gly Ile Ala
    210             215                 220
Ser Leu Val Leu Leu Gly Leu Trp Asp Tyr Leu Asn Glu Ala Ala Ile
225             230                 235                 240
Thr Thr Phe Ser Val Leu Gly Leu Phe Ser Ser Gln Ala Ala Ala Ile
            245                 250                 255
Leu Ser Thr Leu Leu Ala Ala Glu Val Ile Pro Thr Thr Val Arg Gly
        260                 265                 270
Arg Gly Leu Gly Leu Ile Met Ala Leu Gly Ala Leu Gly Gly Leu Ser
    275                 280                 285
Gly Pro Ala Gln Arg Leu His Met Gly His Gly Ala Phe Leu Gln His
    290                 295                 300
Val Val Leu Ala Ala Cys Ala Leu Leu Cys Ile Leu Ser Ile Met Leu
305             310                 315                 320
Leu Pro Glu Thr Lys Arg Lys Leu Leu Pro Glu Val Leu Arg Asp Gly
            325                 330                 335
Glu Leu Cys Arg Arg Pro Ser Leu Leu Arg Gln Pro Pro Thr Arg
            340                 345                 350
Cys Asp His Val Pro Leu Leu Ala Thr Pro Asn Pro Ala Leu
            355                 360                 365
```